Columbia University

Department of Economics
Discussion Paper Series

Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts

Elliott Ash and W. Bentley MacLeod

Discussion Paper No.: 1314-22

Department of Economics
Columbia University
New York, NY 10027

November 2014

Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts

Elliott Ash and W. Bentley MacLeod*

October 28, 2014

Abstract

This paper provides a theoretical and empirical analysis of the intrinsic preferences of state appellate court judges. We construct a panel data set using published decisions from state supreme court cases merged with institutional and biographical information on all (1,700) state supreme court judges for the 50 states of the United States from 1947 to 1994. We exploit variation in the employment conditions of judges over this period of time to measure the effect of these changes on a number of measures of judicial performance. The results are consistent with the hypothesis that judges are intrinsically motivated to provide high-quality decisions, and that at the margin they prefer quality over quantity. When judges face less time pressure, they write more well-researched opinions that are cited more often by later judges. When judges are up for election then performance falls, consistent with the hypothesis that election politics is time-consuming. These effects are strongest when judges have more discretion to select their case portfolio, consistent with psychological theories that posit a negative effect of contingency on motivation (e.g. Deci, 1971). Finally, the intrinsic preference for quality appears to be higher among judges selected by non-partisan elections than among those selected by partisan elections.

*Department of Economics, Columbia University. Contact: eta2103@columbia.edu, bentley.macleod@columbia.edu. We thank Gohar Harutyunyan, Montague Hung, Mithun Kamath, James Kim, Michael Kurish, Justin McNamee, Sourabh Mishra, Arielle Napoli, Bryn Paslawski, Olga Peshko, Quinton Robbins, Carol Shou, Alex Swift, Tom Verderame, Sophie Wilkowske, John Yang, and Fred Zhu for their meticulous help in assembling data and other research assistance. We thank Matthew Freedman, Levon Barseghyan, and Adam Lavecchia for discussant comments on previous drafts. Columbia University's Program for Economic Research and the National Science Foundation Grant SES-1260875 provide financial support for this research.

1 Introduction

This paper provides a theoretical and empirical study of the intrinsic motives of state supreme court judges. We have constructed a panel data set that matches institutional and judge biographical information with published decisions for all state high courts in the United States from 1947 until 1994. We document discrete institutional changes in judicial employment conditions, which either increased or decreased the constraints faced by judges when doing their work. Our within-judge estimates of the behavioral responses to these changes are consistent with the hypothesis that judges are intrinsically motivated to choose higher-quality opinions when given more time, and that they intrinsically prefer quality over quantity at the margin.

This work is important because an independent judiciary is crucial for the operation of civil society. As Epstein et al. (2013) observe, judges are paid on fixed salaries to reduce pecuniary motivation that may bias their decision-making. The first open question is whether and to what extent judges are motivated to provide high-quality decisions – a question which we are able to answer in the affirmative. Second, there is the controversial issue of how judges should be retained and in particular whether they should take an active part in fundraising for re-election (see Spottswood, 2007). We find evidence that in the year a judge is up for re-election their performance falls, consistent with the hypothesis that campaigning takes time. We also find evidence that judges elected via a partisan election process (where judges have party affiliations) have weaker intrinsic motivation to perform than do judges elected via a non-partisan process.

An important source of intrinsic reward to perform is *professionalism* (White, 1959). According to White, professionals are individuals who become personally invested in their skills and are motivated by the challenge of doing a job well. We show theoretically that if judges have a preference for working on influential cases, then constraining their choices can be demotivating and lead to lower performance. This prediction is consistent with work on contingency and motivation developed in the psychology literature (Deci, 1971).[1] Related work in economics includes Frey and Oberholzer-Gee (1997), who provide some evidence that monetary incentives can crowd out an intrinsic motivation to bear the costs of a negative externality to help solve a social problem. The lab experiment results reported in Gneezy

[1]It is also consistent with career concerns models. Dewatripont et al. (1999b), Francois (2000), and Prendergast (2007) show that the conclusions of the simple intrinsic motivation model we develop can also be derived with standard preferences in a more complex dynamic setting. However, concerns about future promotion have limited importance in our setting, where promotion to a higher court (the U.S. Supreme Court) is extremely rare, and almost all judges keep their job until retirement unless they lose election.

and Rustichini (2000) support the notion that intrinsic incentives are blunted by the addition of performance pay.

An open empirical question is the degree to which crowding out effects extend to professionals in their day-to-day work. This question has special importance when a professional is making decisions on behalf of others. For example, physicians are asked to evaluate and recommend treatments in their patients' best interests. An important practical question is whether physicians should receive pecuniary awards for particular treatment decisions, or whether healthcare organizations should trust in physician professionalism and impose flat incentives. Good answers to this type of question have evaded empirical researchers because the healthcare market is complex and dynamic, making it difficult to obtain convincing measures of professionalism and its effect on performance.[2] In this paper, we look at a separate group of highly skilled professionals whose work environment has been stable for many decades: state appellate court judges.

The job of an appellate court judge is to review trial cases and ensure that the law is properly applied. For our purposes this is an attractive setting because the basic functions of the job have not changed much in the last 200 years. Litigants appeal cases where they feel there has been an error. The judges then choose which cases to hear, and then they write a decision regarding the merits of each case. From these written decisions we can build a number of performance measures, such as the number of decisions written in a year, the length of a decision, and how often a decision is cited by later judges.[3] We then estimate the effect of changes in employment conditions on these performance measures using judge fixed effects.

To organize our data and interpret our results we extend the legal-realist model in Epstein et al. (2013) to allow for intrinsic rewards from high-quality decision-making.[4] In our model, judges prefer working on important cases that can influence the law in the future. This additional ingredient is sufficient to reproduce Deci's (1971) result in the judicial setting – namely, giving a judge more discretion over the cases they hear increases effort. Besides providing micro foundations for how extrinsic incentives might crowd out intrinsic motivation, the model leads our empirical inquiry toward institutional reforms that gave judges more time to write decisions and more discretion over their work environment.

The first change we study is the introduction of intermediate appellate courts (IAC's),

[2]See Chandra et al. (2012) for a model of professionalism for doctors and a discussion of the evidence.

[3]See Choi et al. (2008) for a discussion of how to measure judicial performance.

[4]Recent work by economists in a legal-realist tradition includes Glaeser et al. (2001), Gennaioli and Shleifer (2007), and Baker and Mezzetti (2012).

which have the effect of reducing the case load of sitting supreme court judges. We measure the impact of the rule change by comparing the performance of individual judges before and after the introduction of an IAC. These judges could respond in various ways to the reduced workload, including working less (choosing more leisure) or working more intensely on the cases still on the docket. We find that judges respond to the introduction of an IAC by focusing on the more interesting cases that are cited more often by later judges. With more time from the reduced caseload, the judges write longer and more well-researched opinions, suggesting an intrinsic motivation for their work.

Another implication of the time allocation model is that if judges are engaged in outside activities that provide income, then increasing income from judging should reduce time pressure and thus be associated with more and/or higher-quality output. In our data we find that wage increases have a positive effect on performance, but only in the states where judges have full discretion over selecting cases for review. In states with less discretion over their case portfolio, we find no effect of wages on performance. This suggests the importance of control over the work environment in the operation of intrinsic incentives.

Next we consider the question of term length, meaning the number of years a judge serves in between elections. An increase in term length reduces the time pressure on the judge arising from the re-appointment process, for example due to electoral campaigning. In our sample, we find that a judge responds to a term length increase with higher-quality judgments and no decline in output as measured by the number of cases or total number of words written.

This effect of term lengths on performance suggests the influence of electoral processes on how judges allocate their time. In our data we observe three types of electoral systems. We have partisan elections (where judges represent different political parties), non-partisan elections (where judges are not allowed to affiliate with political parties), and uncontested elections (where sitting judges are automatically placed on the ballot with no competition). Lim and Snyder (2013) observe that in non-partisan elections, evaluations by state bar associations have a large effect on voting outcomes relative to partisan elections. The quality of the judge, rather than their political affiliation, is most salient for the non-partisan-election voters. In the period of our data, six states moved from non-partisan to uncontested elections, while nine states moved from partisan to uncontested elections. Under uncontested elections, only the worst judges are removed by vote and electoral demands are much weaker.

We find that the move from non-partisan to uncontested elections is associated with an increase in performance as measured by caselaw research and citations, similar to the effects

of establishing an IAC or increasing term lengths. There is also an increase in the number of dissenting opinions and the number of decisions over-ruled by the legislature, suggesting that the judges are acting more independently. In contrast, the move from partisan to uncontested elections is associated with a weak negative effect on performance. This result is consistent with the hypothesis that judges selected in partisan elections have a lower intrinsic preference for the work of judging.

Finally, we explore the dynamic effect of the election process. We compare the performance of a judge in the year they are up for re-election to other years of their tenure. In years that judges are up for an uncontested election we find no effect on performance, consistent with the hypothesis that these elections impose weak incentives. In contrast, in partisan elections we find evidence of a decrease in both output and quality in the election year, consistent with the hypothesis that election-year politics take time.

In the non-partisan system we observe more nuanced election-year effects. The number of majority decisions written declines, as with the partisan system. However, there is no decrease in quality, research, or opinion length. This is consistent with the existence of time constraints where other judges take on more work when a colleague is up for re-election, and also consistent with a higher degree of professionalism among judges selected by non-partisan elections.

In summary, our results are consistent with the hypothesis that judges have intrinsic motivations that interact with extrinsic incentives to determine the allocation of time across judging and other activities. More specifically, our results suggest that judges are intrinsically motivated to choose higher-quality opinions when given more time, and that they intrinsically prefer quality over quantity at the margin. From an institutional standpoint, our results support the contention that judicial elections interfere with good judging, rather than reward it. These findings provide additional empirical support for adding intrinsic motivation to the standard agency model.

The rest of the paper is organized as follows. In the next section we discuss the relevant literature and the contribution of our results to this literature. Section 3 introduces our model of judicial behavior. The institutional context is discussed in Section 4, followed by a description of the data (Section 5). The empirical strategy is outlined in Section 6 and the results are reported in Section 7. Section 8 provides a concluding discussion.

2 Background

This paper contributes to two distinct literatures. The first is the literature on public sector employment, where the concern is that use of low powered incentives may result in less efficient provision of government services (see Francois and Vlassopoulos, 2008). The second is the large literature on judicial behavior, much of it in political science, that is concerned with understanding how judges make decisions.[5] We discuss each of these in turn.

2.1 Public Sector Incentives

A generic feature of compensation contracts for individuals who are being asked to provide expert opinion is that compensation is insensitive to the opinion provided. The cost for these individuals of changing their recommendation is low, so even small rewards for a particular position can lead to a large distortion in decision-making.[6] In 1991 the National Academies commissioned a report on how to improve compensation policy in government (Milkovich and Wigdor, 1991), concluding that the distortion costs of performance pay outweighed any benefits from increased performance. This report contributed to a large literature, beginning with Kerr (1975), on the dysfunctional behavior that incentive pay can create.

The challenge for economics is how to understand performance in professions where incentive pay is weak. Wilensky (1964) explains this as a consequence of "professionalism" where norms of behavior evolve that create intrinsic incentives – individuals work hard to provide performance that is evaluated by their peers as high quality. Kreps (1997) introduced this idea into economics, leading to a number of interesting theoretical developments, including Dewatripont et al. (1999a,b), Besley and Ghatak (2005), and Benabou and Tirole (2002, 2003, 2006). In Dewatripont et al. (1999a,b), the effort incentive stems from career concerns – being rewarded with a better job in future periods. In Benabou and Tirole (2003), extrinsic incentives crowd out intrinsic motivations because they signal that the principal does not trust the agent.

Francois (2000), Prendergast (2007), Alesina and Tabellini (2007, 2008), and Delfgaauw and Dur (2008) apply these ideas to incentives for public officials. These papers are concerned with how to design public employment contracts under a number of assumptions about the self-motivation of agents. The corresponding empirical inquiry is to measure the extent to

[5] See Epstein and Knight (2013) and in particular Epstein et al. (2013) who have a comprehensive review of the literature.

[6] This result is so well known in incentive theory that it is not typically presented as a result. Variants of the result can be traced to D'Aspremont and Gerard-Varet (1979).

which intrinsic incentives lead public servants to modify their behavior as a function of the environment.

A substantial body of work has used behavioral experiments on college students to tease out the effect of intrinsic incentives, beginning with the seminal work of Deci (1971). Economists have extended this work both in the lab (Gneezy and Rustichini, 2000) and in surveys (Lacetera and Macis, 2010). These papers show that explicit performance pay can crowd out intrinsic motivation. Gneezy et al. (2011) provide a review of this growing literature.

The experimental literature has established the existence of intrinsic incentives and illustrated how explicit incentives can crowd out intrinsic incentives. The more serious challenge is to measure this effect in the field using public sector workers. First, one must establish variation in performance of these workers. For example, Coviello et al. (2011) find that the way judges organize their time affects their performance. Currie and MacLeod (2013) find in a large panel of physicians that there is variation in both the quality and nature of medical decisions.

The next step is to explore how the environment influences behavior. Dal Bo et al. (2013) conduct a field experiment in which they randomize salaries for public sector job offers in Mexico. Higher compensation is associated with increased quality of the applicant pool and better self-motivation among the hired workers. Besley and Coate (2003) find that appointed regulators are more sensitive to the issues that policy makers care about, while elected regulators are more pro consumer. These papers illustrate the effect of selection on intrinsic incentives, a central issue in the political science literature on judging (Knight and Epstein, 1996).

In this paper we are concerned with the effects of changes in the environment on acting officials, rather than on the effects of selection. Examples of this approach in looking at public sector employees includes Mas (2006); he studies the effect of arbitration decisions on police performance and finds that favorable outcomes lead to better police performance. Dal Bó and Rossi (2011) find that longer term limits lead to higher effort by Argentine legislators, consistent with our findings on term length. In a field experiment, Banerjee et al. (2014) find that reforms that reduced managerial autonomy in India police stations reduced police effectiveness.

2.2 Judicial Behavior

One strand of the large literature on judging studies how judicial behavior is formed by the selection process. For example: How does compensation affect the ability of judges selected? Klerman and Mahoney (2005) find that in eighteenth-century England, the passage of statutes giving judges higher salaries was associated with abnormally high returns on stock equity. Baker (2008) instruments for differences in federal judge salaries by comparing them to law-partner salaries in the judge's home state. He finds that these salary differences are not associated with judicial effort, as measured by tendency to dissent in controversial cases and time between case filing and case disposition. Choi et al. (2009) use a cross-section of state supreme court judges to measure the effect of salaries on the quality of judges selected, and they find only mixed evidence on whether judges with higher salaries perform better, as measured by the number of opinions written and number of case citations.

A second literature explores the effect of election politics on the preferences of appointed judges. This work builds on Ferejohn's (1986) point that election politics provide an incentive mechanism. Tabarrok and Helland (1999) find that electoral incentives increase tort awards, while Gordon and Huber (2007) find that they lead to harsher criminal sentences. Shepherd (2009a,b) uses a cross-section of state supreme court decisions to show that judicial voting reflects the political preferences of the retention principal, and that judges seeking re-appointment by another branch of government are more likely to favor litigants from that other branch of government. In a separate cross-section of state supreme court decisions, Choi et al. (2010) find that elected judges write more opinions while appointed judges write more heavily cited opinions.

Lim (2013) compares the behavior of elected versus appointed state court trial judges and finds that as a group appointed judges are more homogeneous and tend to make less harsh sentencing decisions. Iaryczower et al. (2013) extend this work to allow for common values and dispersed information among judges, coming to similar conclusions.

The focus of this study is how individual judges modify their behavior as a function of their environment. This is an important question because it helps us understand how to obtain the best performance from sitting judges. To model their behavior we follow the legal-realist approach pioneered by Landes and Posner (1980), and suppose that judges have well-defined preferences and that they make decisions consistent with those preferences. Recent work by economists in a legal-realist tradition includes Glaeser et al. (2001), Gennaioli and Shleifer (2007), and Baker and Mezzetti (2012). Epstein et al. (2013) introduce an elegant rational choice model of judicial decision-making that explicitly allows for judges to enjoy

allocating time to their work. We cannot directly observe this satisfaction, and hence it is difficult to assess its importance. In order to assess the importance of intrinsic motivation we need to understand how intrinsic motivation affects the outcomes we can observe in our dataset.

Our dataset consists of state appellate decisions, while Epstein et al. (2013) focus on the federal courts. The benefit of using data from the states is that we have a greater variety of treatments across states and across time. Subject to the caveats in Bertrand et al. (2004), we can view the U.S. states as a collection of "natural experiments" that allows us to use a quasi-experimental approach to estimate the causal effect of changes in employment conditions on the behavior of individual judges (see Angrist and Pischke, 2009).

3 Judicial Preferences

In this section we introduce a model of how judges allocate their time between different activities as a function of the importance the judges place on these activities. We begin with a version of the model of Epstein et al. (2013, pp. 25-50) and illustrate how different shocks to the environment lead to time reallocation. Next, we drill down and put more structure on intrinsic preferences and derive some testable implications of the theory of professionalism.

Epstein et al.'s (2013) model begins with the observation that the judge's problem, like most incentive problems, is to allocate time across activities. In this case, those activities are leisure, judging cases, and "outside activities" – that is, non-judging activities that increase (pecuniary and/or non-pecuniary) income. Moonlighting as a private-sector arbitrator or providing legal representation is rare for appellate judges, and generally forbidden by rules of judicial conduct. The key example of an outside activity for our purposes is campaigning for re-election. But as discussed in Epstein et al. (2013), outside activities may include other political activities such as fundraising for party affiliates, writing books and journal articles, or guest lecturing at law schools.

It is assumed that the judge allocates time between leisure T_L, judging T_J, and outside activities T_A, subject to the constraint:

$$T_L + T_J + T_A \leq \bar{T}. \tag{3.1}$$

where $\bar{T}$ is the time available for the period in question (week, month, or year, for example). Let $\vec{T} = \{T_L, T_J, T_A\}$ denote the vector of time allocations. We suppose that the utility

function takes a standard Cobb-Douglas form:

$$U\left(\vec{T},\vec{\alpha},\vec{\beta}\right)=U_L\left(T_L,\alpha_L\right)^{\beta_L}U_J\left(T_J,\alpha_J\right)^{\beta_J}U_A\left(T_A,\alpha_A\right)^{\beta_A}$$

where the vector of judge preferences $\vec{\beta}=\{\beta_L,\beta_J,\beta_A\}$ includes leisure preference β_L, intrinsic valuing of work on cases β_J, and taste for outside activities β_A. The parameters $\vec{\alpha}=\{\alpha_L,\alpha_J,\alpha_A\}$ are used to parametrize the effects of the treatments in our data, as discussed further below.

The goal of the model will be to explore the implications of changes in $\vec{\alpha}$ under the hypothesis that the preferences of the judge, $\vec{\beta}$, are fixed over time. The way a judge allocates her time depends on these preferences, which cannot be directly observed. Our approach is to study within-judge changes over time, which allows us to hold preferences fixed. The various interventions will be modeled as variations in the parameters $\alpha_i, i\in\{L,J,A\}$.

We make the follow assumptions regarding preferences:

Definition 1. Preferences satisfy *continuity* and *monotonicity* if the consumption factors $U_i\left(T_i,\alpha_i\right), i\in\{L,J,A\}$ are strictly positive, twice continuously differentiable, and increasing for $T_i,\alpha_i\geq 0$.

The next assumption ensures that a unique solution exists.

Definition 2. Preferences are *convex* if U_i is concave in T_i for $i\in\{L,J,A\}$.

Finally, we need a condition to sign the effect of the exogenous parameter.

Definition 3. The parameter α_i has a *positive effect* at T_i if $\frac{\partial^2 u_i(T_i,\alpha_i)}{\partial T_i\partial\alpha_i}>0$, where $u_i=\log\left(U_i\right)$.

These assumptions ensure that the optimization problem is concave and has a unique solution. The fact that U_i are concave in T_i and the log function is strictly concave implies that the monotonic transformation of utility, $u=\log\left(U\right)$ gives use an equivalent representation of preferences in a linear form that is strictly concave in $\vec{T}$:

$$u\left(\vec{T},\vec{\alpha},\vec{\beta}\right)=\beta_L u_L(T_L,\alpha_L)+\beta_J u_J(T_J,\alpha_J)+\beta_A u_A(T_A,\alpha_A).$$

The judge's time allocation is assumed to be a solution to:

$$\max_{\vec{T}\in\Re^3_+}\ u\left(\vec{T},\vec{\alpha},\vec{\beta}\right), \tag{3.2}$$

subject to the time constraint (3.1). Our assumptions imply:

Proposition 4. *Under the continuity, monotonicity and concavity assumptions, there exists a unique solution to the judge optimization problem,* $\vec{T}^*\left(\vec{\alpha},\vec{\beta}\right) \geq 0$. *When* α_i *has a positive effect for* $\vec{T}^*\left(\vec{\alpha},\vec{\beta}\right) > 0$, *then* $\frac{\partial T_i^*}{\partial \alpha_i} > 0$ *and* $\frac{\partial T_j^*}{\partial \alpha_i} < 0$ *for* $j \neq i$.

The proof of this proposition is in the appendix.

We can gain a bit more insight into these conditions by defining the *price of time*, given by the Lagrange multiplier, μ, for the time constraint (3.1). The first order conditions for time allocation satisfy:

$$\beta_i \frac{\partial u_i}{\partial T_i} = \beta_i \frac{1}{U_i}\frac{\partial U_i}{\partial T_i} = \mu, \forall i. \tag{3.3}$$

Next, let μ_i represent the marginal value (price) of time allocated to activity i:

$$\frac{\partial U_i}{\partial T_i} = \mu_i, \forall i.$$

In the case of leisure, we assume linear utility that does not depend on α_L:

$$U_L\left(T_L\right) = T_L,$$

which means that $\mu_L = 1$. Thus, from (3.3), we have that $\mu = \beta_L/T_L$. The price of time is pinned down by the preference for leisure and the time allocated to leisure. As the time allocated to leisure increases, the price of time decreases, and vice versa.

Our goal is to understand how changes in the environment affect the value of time – and thereby affect the way judges allocate their time to different activities. In the next two subsections we consider the time allocation to cases and outside activities.

3.1 Preference for Judging

In this subsection we consider the sub-problem of how to allocate time to cases. This involves the construction of a representation $U_J\left(T_J,\alpha_J\right)$ for the utility from judging. First, we suppose that there is a continuum of cases indexed by their complexity/legal significance $\gamma \in [0,1]$, with distribution given by $f\left(\gamma\right)$. The index γ plays two roles. First, the judges have discretion to choose which cases to hear, and, second, judges may choose to devote more time to the difficult or more important cases. The level of judge discretion in reviewing cases may vary across jurisdictions.

The total number of possible cases is given by $N = \int_0^1 f(\gamma)d\gamma$, where the scale can be normalized as desired. Time spent on case γ is given by $t(\gamma)$. We suppose that there is a minimal amount of time needed for each case given by $t_0 > 0$. We define $\delta(\gamma) = t(\gamma) - t_0$ as the amount of time above the minimum to deal with a case.

Each case is assumed to have a basic value to the judge of V_0, which includes the motivation to clear a backlog of cases, not to be seen as lazy, etc. The observation that many judges are given discretion over their case portfolio yet choose to write more than the minimum number of opinions justifies the assumption that $V_0 > 0$. In addition, judges may have an intrinsic or reputational motivation to work on cases of legal significance, which we formalize as follows. Formally, the value of case γ is given by:

$$V(t(\gamma), \gamma) = V_0 + \lambda V_1(\delta(\gamma), \gamma).$$

The parameter λ represents the weight that the judge places on the intrinsic legal interest in a case, given by $V_1(\cdot)$. This value can vary across judges and has two interpretations. First, some judges may simply have stronger preferences for quality and get more enjoyment or satisfaction from the legal work on a hard case. Second, there could be institutional rewards to quality, such as administrative review by a merit commission. A major goal of the project is simply to determine if $\lambda > 0$; that is, whether or not there is evidence of *professionalism* (Wilensky, 1964).

We suppose that judges have a stronger intrinsic motivation to work on the legally important cases. To capture this idea it is assumed that for all $\gamma \in [0,1], \delta \geq 0$, V_1 is twice continuously differentiable and that:

$$\begin{aligned} V_1(0,0) &= 0, \\ \lim_{\delta \to 0^+} \frac{\partial V_1(0,\gamma)}{\partial \delta} &= \infty \\ \frac{\partial V_1}{\partial \gamma} &> 0 \\ \frac{\partial^2 V_1}{\partial \delta^2} &< 0 \\ \frac{\partial^2 V_1}{\partial \delta \partial \gamma} &> 0. \end{aligned}$$

The second condition ensures that should a judge choose to hear a case, then it is always optimal to put more than t_0 units of effort into a case ($\delta(\gamma) > 0, \forall \gamma \in [0,1]$). An interior

solution is not necessary, but it saves on having to keep track of corner solutions where the judge chooses $\delta = 0$. The last condition says that the intrinsic incentive to invest time is greater for higher γ – that is, judges care more about the relatively important cases. These assumptions are sufficient to ensure that judges would choose to hear cases with high γ first.

In states with the death penalty, the supreme court is constitutionally required to review death penalty cases. In some jurisdictions, such mandatory review rules apply to some or all types of appealed cases – for example, cases involving constitutional questions, or felony criminal cases. We suppose that the parameter $\alpha_J \in [0,1]$ represents the fraction of cases that a judge has discretion whether to hear or not, while $(1-\alpha_J)$ is the fraction of cases that must be heard.

Let $\gamma \in [0,1]$ be the lower bound to the cases a judge considers under full discretion. The total gain from the intrinsic value of judging is given by:

$$\begin{aligned} V_J(\gamma, \delta(\cdot)) &= \alpha_J \int_{\gamma}^{1} (V_0 + \lambda V_1(\delta(\gamma), \gamma)) f(\gamma) d\gamma \\ &\quad + (1-\alpha_J) \int_{0}^{1} (V_0 + \lambda V_1(\delta(\gamma), \gamma)) f(\gamma) d\gamma. \end{aligned}$$

The first integral represents value from the discretionary cases, while the second integral represents value from the mandatory cases. The value from allocating time T_J to judging is the solution to:

$$U_J(T_J) = \max_{\gamma, \delta(\cdot)} V_J(\gamma, \delta(\cdot)) \tag{3.4}$$

subject to:

$$\alpha_J \int_{\gamma}^{1} (t_0 + \delta(\gamma)) f(\gamma) d\gamma + (1-\alpha_J) \int_{0}^{1} (t_0 + \delta(\gamma)) f(\gamma) d\gamma \leq T_J. \tag{3.5}$$

The inputs to the model are the strength of intrinsic preference for legal work, λ, the distribution of cases, $f(\gamma)$, and the fraction of cases with discretionary review, α_J. A detailed study of the problem is in the appendix, where we show the following.

Proposition 5. *If V_0 is sufficiently large then $\frac{\partial^2 u_J(T_J, \alpha_J)}{\partial T_J \partial \alpha_J} > 0$. In this case an increase in discretion for the judge leads to more time allocated to judging and more time per case. The more important cases have more time allocated to them. Increasing T_J increases the number of cases heard and the time per case.*

This result provides a theoretical explanation for the demotivating effect of contingency discussed by Deci (1971). Adding contingencies to a task increases the cost of time, which in turn leads to less effort in the professional activities that the judge prefers, namely interesting and important legal cases.

3.2 Preference for Outside Activities

Next we consider the allocation of time to outside activities. We interpret outside activities in general terms; they include any activity that has a cost or benefit other than judging. The main example is electoral campaigning and other political activities, which derive from career concerns about re-appointment. They also include pecuniary activities such as writing books.

We begin by defining:

$$U_A\left(T_A, \alpha_A\right) = I_0 + I_A\left(T_A, \alpha_A\right) \geq 0.$$

The term I_0 is the base income from employment as an appellate judge, while $I_A\left(T_A, \alpha_A\right)$ represents all other forms of non-leisure, non-judging rewards. The parameter α_A represents the return to outside activities. In particular, we suppose that

$$\frac{\partial^2 I_A}{\partial T_A \partial \alpha_A} > 0.$$

Next we have

$$\frac{\partial^2 u_A\left(T_A, \alpha_A\right)}{\partial T_A \partial \alpha_A} = \frac{1}{I_0 + I_A(\cdot)}\left(\frac{\partial^2 I_A}{\partial T_A \partial \alpha_A} - \frac{1}{I_0 + I_A(\cdot)}\frac{\partial I_A}{\partial T_A}\frac{\partial I_A}{\partial \alpha_A}\right).$$

For I_0 sufficiently large, the first term in the parentheses is greater than the second, meaning this cross-partial is positive. Then by proposition (4) we have:

$$\begin{aligned} \frac{\partial T_A^*}{\partial \alpha_A} &> 0 \\ \frac{\partial T_J^*}{\partial \alpha_A} &< 0. \end{aligned}$$

Thus, an increase in the return to outside activities should lead to less time per case and fewer cases heard. The key example of an increase in α_A in our data is being up for re-election, which makes campaigning activity more valuable. Conversely, reduced electoral

incentives due to judge-tenure reforms would reduce campaign-related returns α_A.

Next, we note that the base income from judicial employment, I_0, can also have an effect on performance:

$$\frac{\partial^2 u_A}{\partial T_A \partial I_0} = -\frac{1}{(I_0 + I_A(\cdot))^2}\left(\frac{\partial I_A}{\partial T_A}\frac{\partial I_A}{\partial I_0}\right) < 0.$$

An increase in base income reduces the marginal benefit to time spent on (pecuniary) outside activities. This should result in less time on outside activities and more time on judging (and leisure). However, this only occurs if $T_A > 0$, that is, for judges that are currently engaged in outside activities that produce pecuniary benefits. If $T_A = 0$ then an increase in income will have no effect.

These results are summarized in the following proposition.

Proposition 6. *With stronger electoral concerns (e.g., being in a re-election year), there should be a fall in total output. If a judge is engaged in outside activities for pecuniary returns, then an increase in salary leads to more output and higher-quality decision-making.*

In summary, this model builds on the idea that judges allocate time to different activities in proportion to the return from these activities. If they gain positive reward from allocating time to interesting cases, then any relaxation of the time budget constraint leads to more time allocated to cases, and hence should result in higher-quality decisions. These ideas motivate and organize our empirical work.

The model illustrates that the demotivating effect of contingencies and financial rewards can be explained as a problem in time management. Contingencies increase the value of time in other activities, and hence reduce the time that individuals allocate to activities they prefer. In the case of an appellate judge, constraints on case selection raise the cost of time and reduce the allocation of effort to important cases. These constraints may have a negative effect on total performance.

4 Institutional Background

We study state supreme courts. These courts operate as the state judiciary's analogue to the U.S. Supreme Court, where judges rule on questions of state law rather than federal law. These questions arise in cases appealed from lower state courts. The state supreme court consists of a panel of between five and nine judges, who decide together how to rule on appeals.

A case begins when a plaintiff files a lawsuit or a prosecutor indicts a criminal. At trial, facts are litigated and a judge/jury gives a verdict, which the losing party can appeal. If the state has an intermediate appeals court, they will then take the case and may affirm, reverse, or modify the trial verdict. After this intermediate court's decision (or after the trial decision when the state does not have an intermediate appellate court), the ruling can be appealed to the state supreme court.

If the supreme court accepts a case for review, the panel of judges will rehear the case at oral argument and review the submitted briefs for legal error. Each judge votes whether to affirm or reverse the lower decision. One of the majority judges writes an opinion explaining the decision. In rare cases, the state supreme court ruling is appealed to the U.S. Supreme Court.

This is the institutional context in which we study judicial incentives. Importantly, the job of a supreme court judge does not change much over the course of the career. A judge in his first year of work has essentially the same task as a judge in his last. While the state supreme court judge's job of reviewing cases and establishing precedent is similar to that of a U.S. Supreme Court justice, there are important institutional differences between state supreme courts and the U.S. Supreme Court that provide attractive opportunities for the empirical study of how employment conditions affect judicial behavior.

[TABLE 1.1 HERE]

We identify the causal relationship between employment conditions and judge performance using state laws that reform judicial employment conditions. Our list of reforms is described in Table 1.1. One of the most common reforms is the establishment of an intermediate appellate court (IAC). Intermediate appellate courts significantly filter the set of cases that supreme court judges have to review. When an intermediate appellate court is operating, supreme court judges have a lot of help in reviewing cases and have more discretion in whether to accept cases for review. We expect that the introduction of an intermediate appellate court will increase the time and discretion available to judges, so they should devote more time to what they care about. Following the modeling framework, we conceptualize the IAC treatment as an increase in discretion α_J, which should result in more time spent on judging by Proposition 5.

Next we have monthly data on individual judge salaries between 1974 and 1994. We add to existing findings, such as Choi et al. (2009) and Baker (2008), by measuring the within-judge effect in a panel data framework. This gives us a rich set of variations; as in Mas

(2006), these discrete compensation changes may result in measurable performance changes. Following Proposition 6, we predict that increasing base income I_0 reduces incentives to engage in outside activities and therefore may increase time spent on judging work.

Because compensation is not contingent on performance and because impeachment is rare, the retention process is probably the strongest incentive system facing state appellate judges. Some state supreme court judges have tenure and some have to face election every few years. In the states with elections, we observe changes in the length of term of office. The most straight forward way that we look at the effect of retention incentives is to examine the effects of changes in this term length. In our sample of years, seven states increased the term length, while three states decreased the term length. Increasing term lengths reduces the frequency with which a judge faces re-election, so the effects of election on a judge's time allocation will be reduced. Decreasing term lengths should have the opposite effect. In our model, decreasing electoral incentives corresponds to a decrease in the return to outside (campaigning) activities α_0, which by Proposition 6 should increase time spent on cases and be reflected in higher quantity and/or quality of output.

Next we have discrete rule changes on how a judge is retained. We observe six types of retention systems in our data. The three less common retention systems are those that do not feature judicial elections. In governor retention (five states), the governor decides whether a judge should be re-appointed at the end of his/her term. With legislative retention (three states), the state legislature decides by majority vote whether a judge should be re-appointed. With life tenure (four states), judges cannot be removed except by impeachment. U.S. Supreme Court Justices have life tenure.

We focus on the three most common judge retention systems, which involve elections. In partisan elections (22 states), incumbent judges face a challenger, with party affiliations on the ballot. There are generally two candidates, a Republican and a Democrat, and the incumbent rarely faces a credible primary challenge. In non-partisan elections (18 states), incumbent judges face a challenger, but party affiliations are not on the ballot. There are generally one or two candidates, and the incumbent is not identified as such. In uncontested elections (19 states), incumbent judges face an up-or-down retention vote with no challenger.[7]

The different election systems impose different electoral incentives. One important mechanism for these incentives is voter behavior. Direct evidence of different vote behavior across systems is provided by Lim and Snyder (2013), who look at whether voting is correlated with

[7]Note that in these counts, we include a state in two categories if it changed systems during our panel period.

state bar association evaluations of judge performance. They find that in partisan elections, evaluations are uncorrelated with voting due to party-line voting. In non-partisan elections, evaluations are highly correlated with vote share and probability of winning. In uncontested elections, evaluations are correlated with vote share but not probability of winning; unchallenged judges have de facto tenure. These findings suggest that judges in non-partisan systems have the strongest incentive to spend time impressing voters (and/or bar association evaluators). In partisan and uncontested systems, that incentive is weaker. However, judges in partisan systems also have the demands imposed on them by party organizations, which could also take away time from judging, especially during campaign season.

In our sample of years, six states moved from non-partisan contested election of judges to uncontested election of judges, and nine states moved from partisan contested election of judges to uncontested election of judges. By measuring performance before and after these changes in tenure status, we can assess whether competitive elections incentivize higher judging effort, or whether they instead divert effort away from judging. If judges choose high performance to impress voters or the party organization, then weakening electoral incentives should reduce judge performance. If instead judges have an intrinsic motivation to choose high performance and elections take up their time, then weakening electoral incentives should improve judge performance. This would correspond in our modeling framework to a decrease in the value to campaigning (α_A) after moving to an uncontested system, meaning an increase in judge performance from Proposition 6. The findings in Lim and Snyder (2013) suggest that moving from non-partisan to uncontested elections may have a stronger effect on judge behavior than moving from partisan to uncontested elections.

Another institutional feature that we will find useful is the judicial electoral cycle. Like U.S. Senators, nontenured state supreme court judges face election on a staggered basis, where a subset of judges are up for election in any particular election year. We can compare the performance of judges who are up for election to their colleagues who are not up for election. In our model, being up for election corresponds to a higher return to campaigning α_A and thereby a reduction in judicial output (again, from Proposition 6).

[TABLE 1.2 HERE]

Besides these institutional changes, there are two key rules of appellate procedure that will play a role in our empirical analysis. In particular, we consider the rules for how cases are selected for review by the state supreme court, and the rules for how cases are assigned to authoring judges once they are accepted for review. These rules are listed in Table 1.2.

As emphasized in the theory, the judge's discretion to select cases for review is crucial to the operation of extrinsic incentives and intrinsic motivation. The central idea is that with discretionary case selection, intrinsic motivation should play a more important role in performance because judges have more control over their work environment. We divide state supreme courts into three categories, depending on whether they have full discretion in case selection, partial discretion (that is, some cases are mandatory), or fully mandatory review (that is, all cases must be reviewed to some degree). The U.S. Supreme Court has fully discretionary review. A line of papers in the political science literature have shown that while judges have some level of discretion under mandatory review, it nonetheless has substantial compositional effects on the state supreme court caseload (e.g. Eisenberg and Miller, 2009).[8]

We don't have variation over time within state in the rules for case selection. We study their importance by examining the relative magnitude and significance of our other treatments depending on the case selection rule. In particular, the establishment of an intermediate appellate court should have a larger treatment effect under discretionary case selection, as this gives the supreme court more scope to reduce their caseload.

In terms of the model, this can be thought of as a greater increase in the parameter α_J due to establishing an IAC under full discretionary relative to partial discretion or mandatory review. Similarly, for courts with higher review discretion, treatments that reduce the return to outside activities (α_A or increase base income I_0) might have a larger positive effect on judging effort (relative to courts with lower review discretion) since the return to judging time (relative to leisure time) is greater in those courts.

A second procedural rule that we will find useful is the method of case assignment. At the U.S. Supreme Court, the Chief Justice assigns cases to authoring judges at his own discretion. At state supreme courts, however, this is the minority rule followed in just 15 states. In 13 states, cases are randomly assigned to authoring judges. In the remaining 22 states, cases are assigned on a rotating system, with cases arbitrarily assigned to judges based on their order on the docket. Christensen et al. (2012) examine these rules for a sample of state supreme

[8]In practice, the appellate review standards are relatively complex. Splitting the states into three categories required simplification and some subjective coding decisions. The results on changes in employment conditions without accounting for review discretion can be summarized as follows. First, the IAC effects on research and quality are still positive but lose statistical significance. The salary effects shrink and lose statistical significance. The effects from term length are weaker, but there are still positive and significant effects on case quality. The effects from the non-partisan-to-uncontested reform are qualitatively the same. The effects from partisan-to-uncontested are no longer negative – there are no effects from this reform in the aggregate. The baseline effects of the electoral cycle are mostly unaffected, except that the quality reductions in partisan systems are smaller.

court decisions, showing that with discretionary case assignment, case characteristics are significantly correlated with judge characteristics. This means that when comparing the performance of judges within the same court to each other, estimates from the discretionary system will likely be biased. For random assignment and rotating assignment, however, case characteristics and judge characteristics are only negligibly correlated in Christensen et al.'s (2012) sample. This means that comparing the performance of judges within the same court to each other would be empirically valid with random or rotating case assignment. For the analysis of discrete changes in employment conditions, the case assignment rule does not play a role because the treatment impacts all judges simultaneously. But for the analysis of the electoral cycle, we exploit the staggered election cycle and compare judges to their colleagues using a state-year fixed effect, meaning that restricting our analysis to the random and rotating systems is important.

We have data on several other institutional changes which are not the focus of the present paper but are included in the regressions as control variables. Probably the most important of these is the handful of states that moved from partisan to non-partisan elections. We also have controls for the establishment of a court administrative office, the establishment of a mandatory retirement age, changes in the number of judges, when New York moved from partisan elections to governor retention, and when Pennsylvania moved from single-terms to uncontested re-election.

5 Data

One contribution of this project is the assembly of a new integrated data set on state appellate courts. At present, there is extensive data on federal court judges (e.g., Epstein et al., 2013), but no existing comprehensive panel data on state courts. Existing studies, such as Landes and Posner (1980), Shepherd (2009b), Tabarrok and Helland (1999), Hall and Bonneau (2006), Gordon and Huber (2007), Lim (2013), and Iaryczower et al. (2013), ask different questions and/or use different, shorter time periods. Our data set has a much longer time period and is more comprehensive. The State Court Data Project had four years of data (1995-1998), and did not have data on how often cases were cited by later judges. Choi et al's (2009, 2010) data included three years of cases (1998-2000). These short time frames only allow cross-sectional studies of judge behavior. Since our data set spans 48 years of data (1947-1994) it allows us to use a within-judge identification strategy.

We have constructed three data sets, with three types of data: 1) judge characteristics,

2) institutional variables, and 3) judicial output measures. We discuss these sets of data in the next three subsections.

5.1 Judge Characteristics

First we have data on the characteristics of individual judges. A team of research assistants collected these data from a range of sources and built biographies for each judge in the sample. The key sources include state court web sites, judge obituaries, and Marquis Who's Who. Items that were unavailable from these sources were obtained through interviews of state court administration staff.

[TABLE 2.1 HERE]

Table 2.1 presents summary means and standard deviations for a collection of judge variables in our sample. We present separate statistics for the different processes by which the judges were selected. State supreme court judges are hired in middle age: 53.6 years old on average. They work as judges for an average of 12 years, are overwhelmingly male, and most of them resign or retire (rather than earn a promotion, die in office, or lose election). Because promotion (defined in this table as moving to a federal appeals court or to a governorship) is so rare, career concerns are likely a limited source of incentive pressure for these judges. We don't see any large differences across the electoral systems.

5.2 Institutional Variables

The key changes in institutional treatment variables are listed in Tables 1.1 and 1.2. They are described in Section 4 above. Summary statistics for these variables by state and year are listed in Table 2.2. The discrete rule changes are represented in the data as dummy variables that equal one for the years after the law change and zero for the years before. In almost all cases, reforms are enacted by voters through ballot referendums administered in November and officially going into effect the subsequent January. In these cases the dummy variable would turn on in the year following the vote. In cases where the policy is effective in the first half of the year, it is coded as turning on in that year. For annual salaries, we give the weighted average across months. The term length changes are combined in a single variable, where a term length decrease is represented as a negative one. When Tennessee moved back to a partisan system in 1974, that is coded as a negative one in the partisan-to-uncontested

treatment variable.[9]

[TABLE 2.2 HERE]

The institutional variables were collected from a range of sources. Most of the judge salary data was obtained from the National Center for State Courts (NCSC), which administers an annual survey of state judge salaries. These data were error-checked by research assistants. We use BLS data on prices for regions or MSA's, when applicable, to adjust for inflation. The appellate review standards were also collected from the NCSC. The case assignment rules were taken from Christensen et al. (2012).

The data on discrete rule changes were collected from previous papers on judge election rules, from the web site judicialselection.us, and from inspection of legislation and constitutional amendments. Hanssen (2004) previously coded the election-system changes, although there were several errors in his data which we have corrected. The other information, for example on term lengths, was collected from the court web sites and other historical sources.

5.3 Judicial Performance

Our third set of data are judicial performance measures constructed for judicial opinions. Here we use performance measures suggested as important by Landes and Posner (1980), Choi et al. (2010), and Epstein et al. (2013). These include the number of opinions written, the length of opinions, the amount of research put into opinions, and the number of subsequent citations to a judge's opinions. These data were collected by a team of research assistants from Bloomberg Law (bloomberglaw.com), which has the items we need presented in a standard format that is amenable to automated parsing. We load the case text and institutional variables into a PostgreSQL relational database, which is ideally suited to the simultaneous analysis of large text corpora and structured numerical data. A series of Python scripts interface with the Postgres database to parse the text, compute performance measures, and merge with the institutional variables. We export to CSV files for use in Stata.

Our data is constructed from the universe of opinions published by state supreme courts between 1947 and 1994. The full sample includes 1,025,461 cases. Many of these cases are summary orders – certiorari denials, habeas corpus denials, and other brief orders that do not require a full written opinion. These orders are just a few sentences long and are rarely if

[9]Treating these changes symmetrically is a strong assumption, but most of our estimates are similar if we treat them as separate reforms. Statistical significance is lower due to the smaller number of experiments.

ever cited by future judges. Many states in our sample do not publish these types of orders. Our interest is in the professionally authored legal precedents that explain the ruling for future judges, so we exclude summary orders from our empirical analysis. Specifically, we focus on published majority opinions that are seven or more sentences in length – orders with six or fewer sentences are removed. This step shrinks the sample to 496,099 majority opinions. Next, we are interested in the behavior of individual judges across time. We therefore remove unauthored (per curiam) opinions, as well as the small number of opinions written by non-supreme-court judges, such as magistrates, commissioners, and other special sitting judges.[10] This step shrinks the sample to 387,905 majority opinions (plus attached discretionary opinions) written by judges for which we have biographical information. This divides down to 184.7 cases per state per year and 25 cases per judge per year on average.

The list of performance variables, along with summary statistics, are presented in Tables 2.3 through 2.7. First, we use the number of majority and discretionary opinions written. At the state supreme court level, whether to accept a case for review is often discretionary, so if judges accept more cases for review they are taking on more work. Whether to write a discretionary opinion—a concurrence or a dissent—is always up to the judge's discretion and involves willingly taking on more work.

Second, we construct effort statistics from the raw text of a judge's opinions. An appellate judge's output is his writing; a rough measure of increased effort would be increased language output. Here we rely on the total number of words written during a time period, as well as a basic opinion length measure – the average number of words per majority opinion written. We also have a measure of the amount of research a judge engages in – the average length of the Table of Cases gives the number of previous authorities cited in her opinions.

When using opinion length and caselaw research as effort measures, it's important to note that a lot of the raw labor inputs into opinion writing and research are provided by supreme court clerks. If we see changes in output in response to rule changes, that effect may be due in part to changes in how the judge manages her clerks. From our interviews with state supreme court court staff, it seems that the number of clerks per judge remained relatively constant over the time period of our study. Moreover, the processes of clerk selection and retention were relatively stable and do not seem to be correlated with changes in other judicial institutions.

[10]Our treatment variables are uncorrelated with the number and proportion of per curiam opinions, with the exception of the establishment of the intermediate appellate court. The proportion of per curiam opinions goes down after an IAC is established. This reflects that per curiam opinions are on average less important than authored opinions, and that after an IAC is established the court reviews fewer less important cases.

Third, we have the number of citations to a judge's opinions by other judges. In our data, Bloomberg Law staff attorneys have categorized citations as positive, distinguishing, or negative. A positive cite is a clear signal that a decision is found useful by a future judge. A distinguishing cite means that part of the ruling is useful, but needs to be clarified – so this is perhaps a weaker signal of opinion quality. The significance of a negative citation is more problematic; the most intuitive interpretation is that a negative cite means a judge made the wrong decision. On the other hand, negative cites could mean that a judge is being more creative in his judging and allowing for more experimentation in lawmaking. A final possibility is that negative cites are just another signal of an opinion's influence relative to other opinions, and therefore could serve as an additional quality measure.

Using Bloomberg's citation analysis features, we can construct more fine-tuned data on judicial citations. We have information about whether a case is discussed by the future court (rather than cited without comment) and whether it is directly quoted by the citing court. These measures can be understood as more direct signals that the citing court finds the opinion useful. The Out-of-State Cites measure includes positive cites from out-of-state courts; as noted by Choi et al. (2010) among others, this is perhaps the best quality measure because the cited case serves as persuasive rather than binding precedent. For all of these citation measures, however, an important caveat is that the number of cites is a joint measure of both the importance of a case and the effort of the judge.

Finally, we have the number of cases overruled by later courts and the number of cases superseded by statute (that is, overruled by the legislature). Higher scores on these measures might be seen as poor judging, since other judges or branches of government are reversing their decision. On the other hand, higher scores on these measures might also be interpreted as a sign of greater judicial independence, which as mentioned could lead to more experimentation in lawmaking.

[TABLE 2.3 HERE]

[TABLE 2.4 HERE]

Table 2.3 provides summary statistics for case-level variables. Besides the performance-related variables just described, this table includes a handful of opinion-level variables that are not used in the empirical analysis but provide context for the type of work performed by state supreme court judges. We have the case outcome – affirm, reverse, remand, or modify – and the main areas of law for each case – civil, criminal, administrative, or constitutional. The first pair of columns gives statistics for the full sample of cases, while the second pair of

columns gives statistics for the pruned sample of authored legal opinions, which is the sample of opinions used in the empirical analysis. As expected, the average opinion in the pruned sample is longer, more well-researched, more well-cited, and has more discretionary opinions attached. This reflects that the less important summary orders have been excluded. Table 2.4 gives summary correlations within case for the set of performance measures used in the empirical analysis. Notably, all of our measures are positively correlated within opinion, some strongly so. In particular, the fact that negative cites are strongly correlated with positive cites would support the use of negative cites as an additional signal of opinion quality.

It's tempting to draw causal inferences from these correlations; after all, a well-researched opinion is likely to be more well-cited due to its higher-quality research. But there are other unobserved qualities of opinions that also influence the number of citations, such as the clarity of the legal reasoning or the novelty of the legal issues presented, which are likely also correlated with the length of the table of cases. While it is an interesting open empirical question what case features determine the number of citations, we do not have independent exogenous variation in particular features of opinions and therefore cannot procure credible estimates of those effects. Therefore in our empirical analysis we use each performance variable separately and only as a dependent variable. We remain agnostic about the causal relationships between the various performance measures.

[TABLE 2.5 HERE]

Table 2.5 aggregates the data by judge year. This level of aggregation is used in the results reported below because it allows the use of a judge fixed effect and treats a year of work by an individual judge as the unit of observation. If we used case-level data in the regressions, then the number of opinions written would skew a judge's weighting in the estimates. Also, it makes sense for judges who work many years to count more than judges who work just a few years. Finally, using years rather than months or quarters is helpful because we avoid problems associated with seasonal variation in performance, for example due to vacation time.

The statistical levels observed in the data are not especially relevant to our empirical analysis. Our coefficient estimates are derived from log specifications and can therefore be interpreted as proportional changes due to the treatments. That said, one might note that each of these judges is responsible for a large corpus of output in any given year. The average number of words written annually, 63,831, is the length of a short novel. On average, a judge's opinions for a year are used 43 times by judges in other jurisdictions (Total Out-

of-State Cites), illustrating that state supreme court judges are influential figures that can play an important role in the broader legal system.

[TABLE 2.6 HERE]

Table 2.6 reports judge-year summary statistics with separate columns for the three election systems. States with governor re-appointment, legislative re-appointment, or life tenure are not included in this table. The election systems are similar on most measures. In the uncontested merit systems, judges write longer opinions, and they have longer tables of cases (caselaw research). They are also superseded by statute more often. In the non-partisan systems, citations per opinion are lower in some categories. These differences in levels of the performance variables across systems and across judges could be due to differences in institutional variables, to selection of different judges, or just to unobserved cultural factors or norms. We cannot say that a specific measure on a particular performance variable indicates high performance or high quality. Instead, we look at proportional changes in the variables relative to a judge's baseline in response to treatment, making the more modest assumption that higher measures on these variables relative to baseline means higher effort or quality for a particular judge.

[TABLE 2.7 HERE]

Finally, Table 2.7 reports summary correlations for the performance variables using the judge-year data set. As with the case-level correlations reported in Table 2.4, the judge-year correlations indicate that the performance measures are mostly correlated within judge over time. The multiple measures of output and quality provide multiple signals of the amount of time a judge spends on his opinions. The important exception in this table is the number of majority opinions, which is negatively correlated with most of the quality measures. This suggests that judges face a tradeoff between quantity and quality. These correlations may reflect that judges with larger caseloads have to sacrifice on quality, or it may just be a case composition effect where judges with larger caseloads also tend to work on less important cases. The goal of our empirical work is to discriminate between these types of explanations using panel data.

6 Empirical Strategy

The core of our empirical approach is to exploit *within-judge* variation in performance. More precisely, consider the following setup with three types of data. First, we have judge char-

acteristics X_i, which include observables such as age, experience, and education, as well as unobservables such as ability and preferences (These would include λ and $\vec{\beta}$ from the model). Second, we have time-varying employment conditions Z_{it}, which generally vary within state over time but could also vary by judge within state (due to a staggered electoral cycle, for example). These variables include compensation, rules for appointment and retention, term length, etc. Third, we have judge performance outcomes Y_{it}, which are constructed from the sample of judicial opinions as described in Section 5.3. We would like to compare judge performance under conditions Z^A and Z^B, such that the causal impact of potential outcome A compared to B is given by:

$$Y_{it}^A - Y_{it}^B = F\left(Z^A, X_i, t\right) - F\left(Z^B, X_i, t\right)$$

where $F(\cdot)$ describes the outcome as a function of the treatment (Z), judge characteristics (X_i), and time (t).

Previous papers on judicial employment conditions have taken two different approaches to solving the identification problem. The structural approach, best-known from the industrial organization literature (e.g., Angrist and Pischke, 2010), assumes that the model of judge behavior is known and that only parameter values are unknown. Two leading applications of this approach to state supreme court judges are Lim (2013) and Iaryczower et al. (2013). These types of papers illuminate the relative importance of different mechanisms assuming that those mechanisms exist, but they cannot demonstrate the existence of causal effects.

A second approach is to estimate cross-sectional effects conditional on observables. The two leading applications of this approach to state supreme court judges are Choi et al. (2009) and Choi et al. (2010). This approach assumes that X_i captures all the relevant characteristics of a judge. Then whenever $X_i = X_j$, we have that

$$F\left(Z, X_i, t\right) = F\left(Z, X_j, t\right), \forall Z.$$

If we observe the Z assigned to different judges, we can estimate the causal effect via

$$Y_{it}^A - Y_{it}^B = F\left(Z^A, X_i, t\right) - F\left(Z^B, X_j, t\right)$$

and no time variation is needed. The problem with this approach is that judges and courts may have unobserved characteristics that vary systematically by state. If so, correlation between performance and employment conditions are not likely to be causal links, but the

result of differences between the judges selected.

Our solution to the identification problem is to use the panel structure in our data. Intuitively, we view the 50 states of the United States as 50 potential experiments. We use changes in state laws determining judicial employment conditions over time as natural experiments, measuring the changes in judicial performance in response to changes in employment conditions. Formally, we are interested in measuring:

$$Y^A - Y^B = \sum_{(i,t)\in T(A,B)} \frac{F\left(Z^A, X_i, t\right) - F\left(Z^B, X_i, t-1\right)}{\#T\left(A,B\right)},$$

where $T\left(A,B\right)$ is the set of all the judges i and periods t where employment rule Z^A prevailed in period t and employment rule Z^B prevailed in period $t-1$. We hold fixed as many state- and judge-level characteristics as possible, with the hope of identifying the causal effect of the change from Z^B to Z^A. Bertrand et al. (2004) show that, if one includes state time trends and time dummies, under relatively weak conditions one can correctly identify the effect of a change at the state level on individuals in that state. A number of studies have used this approach in a law-and-economics context, including Miles (2000), Autor et al. (2004), Autor et al. (2006), MacLeod and Nakavachara (2007), Currie and MacLeod (2008), Carvell et al. (2012) and Avraham et al. (2013).

Our econometric specification is a linear model estimated by ordinary least squares. We index records by ist, where we have judge i, state s, year t.[11] Our set of judge characteristics, described in Section 5.1, is represented by X_i. Our vector of treatment variables, described in Section 5.2, is represented by Z_{ist}; what is included varies by regression and is described in more detail in the results section. We have a set of performance measures Y_{ist}, described in Section 5.3, which are constructed from the sample of opinions written by judge i working in state s during year t. The outcome variable in our regressions is $y_{ist} = \log(1 + Y_{ist})$; coefficients can therefore be interpreted as proportional changes due to reforms.[12]

Formally, we estimate

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + Z'_{ist}\rho + \epsilon_{ist}$$

where TIME_t is a fixed effect for year t, JUDGE_i is a judge fixed effect, and $\text{STATE}_s \times t$ is

[11] More precisely, it is *court* s, since Oklahoma and Texas each have two courts of last resort.

[12] We use $1+Y_{ist}$ to account for zeros in the data; the means of the variables are mostly far from zero. Our results are robust to using levels or the inverse hyperbolic sine rather than logs of the dependent variable. The adjusted R^2 is usually higher in the logs specification than in levels.

a state-level linear time trend for state s. The year fixed effect allows for arbitrary national trends in the performance variable. The judge fixed effect controls for time-invariant state-level and judge-level characteristics. The state time trends control for preexisting trends in the performance variable that may be confounded with changes in state laws.[13]

Consistency requires that, conditional on all other covariates, the treatment variable is uncorrelated with the error term. There are two arguments in favor of conditional independence of rule changes and performance. First, as suggested in Hanssen (2004), amendments to judge employment contracts are proposed as a result of external political conditions, rather than in response to judge performance. Second, a majority of voters (or legislators) have to ratify these amendments, meaning that passage has a random component.[14] These arguments strengthen the notion that changes in employment contracts cause changes in judging effort, rather than vice versa.

Following Bertrand et al. (2004), we use robust standard errors clustered at the state level, allowing for heteroskedasticity and for arbitrary correlation of the error term within state across judges and across time. This is sensible because unobserved shocks to performance are likely correlated within the same court. This clustering method is necessary for valid statistical inference.[15]

7 Results

This section presents our empirical results. We have five sets of results. First, Section 7.1 reports results on establishing an intermediate appellate court. Results on increasing judge salaries are presented in Section 7.2. The results on extending judge term lengths are in Section 7.3. Section 7.4 presents the results on changing electoral systems, while Section 7.5 reports the effect of being up for election in various electoral systems.

[13]Our results are robust to the inclusion of a full set of dummies for years of judge experience, meaning that the effects are not generated by mechanical changes in judge human capital. The results are robust to removing the first and last years of each judge's career, meaning that they are not generated by outliers related to different case compositions for younger/older judges, or for judges transitioning between positions. Finally, when we use a fixed effect for a cohort of judges to control for judge turnover, the coefficient estimates have the same sign but are smaller since many of the effects occur toward the end of the ten-year effect window.

[14]And indeed, we observe several failed reforms that attempted to make the same changes that we examine here. We do not observe the same effects on performance as the successful reforms (but it is a small sample of states).

[15]Our statistical tests are robust to using two-way clustering by state and year. See Cameron et al. (2011).

7.1 Effect of an Intermediate Appellate Court

First we look at the effect of introducing an intermediate appellate court (IAC). When an intermediate appellate court is operating, supreme court judges have a lot of help in reviewing cases and have more discretion in whether to accept cases for review. Reducing the caseload should mechanically reduce time demands for the judges, and hence these regressions provide a robustness check. In terms of the model notation, this reform can be seen as reducing t_0, increasing $\bar{T}$, and/or increasing α_J.

Note that in the case of the IAC reform, the assignment to treatment is clearly not random. These courts were established because the supreme court was overworked, and anecdotal evidence suggests that the supreme court judges actively lobbied for the lower court.[16] We can measure the effect of the treatment on the judges in these states. But we can't make strong external validity claims about the effects of an IAC in states that did not choose to establish one.

Formally, we estimate

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \bar{Z}'_{st}\bar{\rho} + Z'_{st}\rho + \epsilon_{ist} \tag{7.1}$$

where TIME_t is a fixed effect for year t, JUDGE_i is a judge fixed effect, and $\text{STATE}_s \times t$ is a state-level linear time trend for state s. The term $\bar{Z}_{st}$ is a vector of indicators equaling one for the baseline time windows of ten years before and ten years after each of the policy changes discussed in Section 4. Z_{st} is a vector of treatment indicators for the ten years after each rule change. Thus, with the inclusion of the judge fixed effects, the estimates for the elements of ρ can be interpreted as the average difference in within-judge performance for the ten years after the policy change relative to the ten years before the policy change. This ten-year-window specification accommodates the average career length of state supreme court judges – having a longer effect window would give too much weight to the handful of judges who work on the court for many years before and after the reform. Using a shorter effect window weakens our effects – it takes a few years for the full effects of the reforms to materialize.

[TABLE 3.1 HERE]

[16] For example, the Massachusetts judiciary web site states: "The Supreme Judicial Court's appellate caseload had greatly expanded through the late 1950s and 1960s. Expansion was fueled in part by a huge increase in criminal appeals... Supreme Judicial Court Chief Justice Joseph Tauro, with considerable support and assistance from SJC Clerk John E. Powers, leaders of the Legislative and Executive Branches, and the state's bar associations, succeeded in getting an intermediate appellate court established in 1972."

Our estimates of the effect of establishing an intermediate appellate court are reported in Table 3.1. Each coefficient is from a separate regression, with the outcome variable in the leftmost column and the three subsequent columns including the three estimating specifications, using different fixed effects and trends. Column 1 estimates the intermediate appellate court effect by comparing across states within years, basically giving the average difference in supreme court judge performance between states that have an intermediate appellate court and those that don't. Column 2 gives the within-state effect, which compares how the supreme court as a whole did after establishing the IAC, with both sitting judges and newly arrived judges. Column 3 gives the within-judge effect, looking only at the average treatment effect on sitting judges at the time of the rule change.

As expected, adding an intermediate appellate court reduces the number of opinions written. After the change the supreme court judges are sharing the caseload with a lower court. Interestingly, the decrease in opinions written does not significantly decrease the total number of words written because judges are compensating by writing longer majority opinions. On average, the cases are more well-researched. The longer, more well-researched opinions are also of higher quality, as measured by positive cites, distinguishing cites, discuss cites, quoted cites, and out-of-state cites. The increase in negative cites may reflect a greater level of judicial independence and experimentation, which is consistent with their having more time and energy in opinion-writing.

As emphasized in the theory, the level of discretion given to state supreme court judges in selecting cases for review should matter a lot in how they reallocate their time in response to changes in employment conditions. Therefore we have interacted the effects of the treatments with the level of discretion given in each state. Formally, we have estimated

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \bar{Z}'_{st}\bar{\rho} + Z'_{st}\rho + \gamma^{PD} Z'_{st}\rho^{PD} + \gamma^{MR} Z'_{st}\rho^{MR} + \epsilon_{ist} \quad (7.2)$$

where $\gamma^{PD}$is a dummy equaling one in the partial-discretion states, while γ^{MR} is a dummy equaling one in the mandatory-review states (as described in Section 4 and listed in Table 1.2). These dummies were included as controls in the previous table – and they were significant in the sense that model fit improved with their inclusion. The coefficient ρ, reported in Table 3.1, measures the baseline effect of establishing an intermediate appellate court in the full-discretion system, while the coefficients ρ^{PD} and ρ^{MR} respectively give the additional interacted effects under partial discretion and mandatory review. Note that the differences in results between the case selection systems do not have a causal interpretation. While the

theory gives us good reason to believe that discretion matters, these rules do not vary over time and could be confounded with other court and judge characteristics.

[TABLE 3.2 HERE]

Table 3.2 reports the baseline within-judge effects from Table 3.1 Column 3 along with the interacted effects by appellate review standard. Note that the number-of-opinions coefficients are positive – they are not significant, but the difference is proportionally large, consistent with lower discretion constraining the ability of judges to reduce their caseload in response to the reform. In turn, some of the effects on quality are reduced – relative to the baseline effect with full discretion, the IAC effect in partial-discretion states is weaker (that is, negative) for opinion length, research, and opinion quality. For opinion length, positive cites per opinion, and discuss cites per opinion, that difference is statistically significant. For mandatory review, we don't see much difference compared to full discretion, but there are only three states in this sample (few states keep fully mandatory review if they have an intermediate court).[17]

The effects on per-opinion performance measures in Tables 3.1 and 3.2 are potentially from two sources. First, an IAC increases judge discretion over case selection, so they may be selecting a set of cases that are more interesting on average. Second, judges may be putting more effort into the cases left over. The average length, research, and cites might increase without increases in judge effort just because the set of cases is different. To study this, we look at how our performance measures change on a fixed number of the most important cases in a judge's portfolio, with the idea that higher performance on these cases is due to higher effort in response to having more time. Specifically, we construct a set of judge-year performance data using averages from the five lowest-quality opinions and five highest-quality opinions published for each judge, ranked by the number of positive citations. This necessitates the exclusion of a handful of judge-years with fewer than ten opinions. If

[17]For completeness we also ran regressions interacting the establishment of an intermediate appellate court with the type of retention system in the state. The decrease in number of opinions after an IAC is established is concentrated strongly in the partisan election states. Partisan judges are also the only ones that reduce their total number of words written. In light of our other evidence, this could be interpreted as suggesting that these judges care less about their work and have the strongest desire to reduce their workload when given more discretion over it. Average majority opinion length and length of the table of cases increase in all of the systems, although the standard errors are larger in the non-partisan and uncontested systems. The increase in caselaw research per opinion is actually largest for partisan judges – but likely reflects that they are no longer reviewing the less important opinions. Among the elected judges, the increase in positive citations and quoted-in citations are only seen in the non-partisan system. In the partisan system and uncontested system, there are no effects on these measures. There are mixed results for the other quality measures.

we observe a change only in the bottom-quality cases, that suggests the IAC effect consists solely of a change in the composition of cases. If we observe a change in the top-quality cases as well, that suggests the judges are using their extra time to put more work into important cases.

[TABLE 3.3 HERE]

The effect of establishing an intermediate appellate court on the bottom five and top five cases by quality are reported in Table 3.3. As with Table 3.2, we report the interacted effects by appellate review standard. As expected, the bottom 5 cases (Column 1) show large increases in most quality measures, reflecting that the less important cases are no longer being accepted for review and the tail of the distribution is being cut off. In terms of our model, this is consistent with the analysis of λ and $V_1(\cdot)$ where judges rule on the more important cases first.

The effects on the top 5 cases (Column 2) are weaker, with evidence of higher effort consisting of longer opinions with longer tables of cases. This suggests that with the extra time from a reduced caseload, judges are spending more time on their most important cases, writing longer majority opinions that are more well-researched. For this reform, however, that extra work does not translate into more citations.

Under partial discretion (Columns 3 and 4), meanwhile, these positive effects are significantly weakened, for both bottom-end and top-end cases. Relative to the full discretion case, these judges do not increase effort as much because their workload is not reduced as much. For mandatory review (Columns 5 and 6), the effects on the bottom-end cases are reversed on a few measures. At the top end, we actually see improvements in quality for the mandatory review states, suggesting that with what extra time they get from the reform they spend it on their most important cases.

Overall, these results are consistent with the hypothesis that judges have an intrinsic incentive to do good work. When given more discretion to select cases, judges choose the more interesting cases that have a stronger impact on the law. When their workload is reduced due to a smaller caseload, they spend more time working on the cases remaining on the docket.

7.2 Effects of Judge Salary Changes

In this section we measure the effect of judge salary on judge performance. This analysis is motivated by Proposition 6, which states that increasing base income I_0 should reduce time

spent on outside activities and thereby increase time spent on judging. We use a standard panel data setup with annual judge performance data. Specifically, the outcome y_{ist} for judge i in state s at year t is modeled as

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \rho Z_{st} + X'_{st}\beta + \epsilon_{ist} \tag{7.3}$$

where TIME_t is a fixed effect for year t, JUDGE_i is a judge fixed effect, and $\text{STATE}_s \times t$ is a state-level linear time trend for state s, Z_{st} is the log real annual salary paid to judges in state s at year t, and ϵ_{ist} is an error term. The set of control variables X_{st} includes treatment dummies for all of the institutional reforms listed in Table 1.1 and described in Section 4. This regression effectively compares deviations from the detrended mean log salary to deviations from the detrended outcome variable. Outcome variables are in logs, so coefficients can be interpreted as the predicted percent change in the outcome variable for a one percent increase in real salary.

[TABLE 4.1 HERE]

Table 4.1 reports the salary results. Column 1, with only year fixed effects, is the closest specification to that used in Choi et al. (2009). Column 2 gives the within-state effect which would include the effect of salary increases on sitting judges and the quality of new judges. For analyzing only the incentive effect, we look at Column 3.

Consider first Column 1, which compares judges across states. There are large and significant effects. Judges with higher salaries tend to write fewer opinions, but they are longer and have more citations. In Columns 2 and 3, the effects mostly disappear. This reflects the unobserved heterogeneity in court rules and in judge characteristics across states, which are correlated with both salaries and our performance measures. Due to other judge-level and court-level factors, those judges paid higher salaries are also the ones who write fewer, better opinions.

However, we do see weak evidence of long-term incentive effects of salary changes. In particular, both within-state and within-judge, increases in log salary are associated with increases in positive cites, discuss cites, and quoted cites per opinion. The other measures are positive as well, although not statistically significant. The within-judge measures are pretty much identical to the within-state measures, meaning that these are incentive effects on sitting judges.[18]

[18]Interacting salary changes with the retention system shows that there are similar (small positive) effects of log salary on effort across the different electoral systems.

As done in Section 7.1 with intermediate appellate courts, we have interacted the effects of salary changes with the level of review discretion in each state. Specifically, we have estimated

$$y_{ist} = \text{TIME}_t + \text{JUDGE}_i + \text{STATE}_s \times t + \rho Z_{st} + \rho^{PD}\gamma^{PD} Z_{st} + \rho^{MR}\gamma^{MR} Z_{st} + X'_{st}\beta + \epsilon_{ist} \quad (7.4)$$

where as before, $\gamma^{PD}$is a dummy equaling one in the partial-discretion states, while γ^{MR} is a dummy equaling one in the mandatory-review states. These dummies were included in the Table 4.1 regressions, which reported the baseline salary effect for full-discretion states.

[TABLE 4.2 HERE]

Table 4.2 highlights the importance of discretionary review in this result. As Table 3.2 did with IAC's, this table shows the baseline within-judge effect as well as the interacted effects with partial discretion and mandatory review. While we see salary-related increases in opinion quality in the full-discretion states, the judges in mandatory-review states respond to salary increases by decreasing quality. The effect is statistically negative if you look at these states individually without interacting them with the full-discretion states.

The effects on time use of relieving time pressure from outside activities is stronger when a judge has discretion over using his time. One way to interpret this result is that time spent on judging is less valuable under mandatory review. They have less intrinsic motivation for their job due to having less control over their work environment, so the leisure effect is stronger than the intrinsic-motivation effect. We emphasize again, however, that these differences across the case-selection systems could be due to selection of different judges and have nothing to do with the incentive effects of the discretion rules.

7.3 Effects of Term Length Changes

Here we estimate the effects of changes in term lengths, as discussed in Section 4. With a longer (shorter) term of office, judges face weaker (stronger) electoral incentives because they have to face election less (more) often. Therefore increasing the term of office should result in judges spending more time on what they care about.

This rule change occurred in ten states in our sample, but it is independently identified in only eight states due to the co-occurrence of other reforms in two of the states. Of these states, at the time of the reform Hawaii had governor retention, Vermont had legislative retention, Kentucky, Montana, and South Dakota had non-partisan elections, Louisiana

had partisan elections, and Illinois, Indiana, Pennsylvania, and Maryland had uncontested elections.[19]

[TABLE 5.1 HERE]

The results reported in Table 5.1 come from estimating Equation (7.2). As with the IAC effect in Section 7.1, Column 1 gives the across-state effect, Column 2 the within-state effect, and Column 3 the within-judge effect. The Column 3 coefficients capture the average effect of a term length change on all judges who are active at the time the change occurred in each state relative to the state-specific trend, in the ten years after the policy change, relative to the ten years before the policy change. As with the IAC regression, dummies for all other rule changes are included.

As can be seen in Column 3, when judges get more time due to changes in term length, they respond by increasing opinion quality. While there are no significant effects on number of opinions and opinion length, there is an increase in caselaw research and an increase in several of our opinion quality measures. After an increase in the term length, we see improvements on positive cites, distinguishing cites, discuss cites, quoted cites, and out-of-state cites. As with the IAC, there is also an increase in negative cites, suggesting that they are writing opinions that matter, but which are controversial. These estimates are consistent with the hypothesis that longer terms make intrinsic incentives more powerful, with the result that judges invest more in the task of opinion writing.

[TABLE 5.2 HERE]

To further explore the importance of discretion over case selection as emphasized in the model, Table 5.2 gives the baseline within-judge effects from Table 5.1 along with the interacted effects by appellate review standard. With this reform, the limitations of partial discretion in case selection do not seem to restrict the salutary effect, with these two states actually demonstrating stronger effort/quality responses in some ways. With the two mandatory-review states, however, notice the following: The term-length effects on caselaw research and opinion quality are reversed. They are statistically significantly different, in

[19]Interacting the term length change with the retention system results in the following observations. The positive effects on quality are strongest in the governor retention, non-partisan election, and uncontested election systems. Unlike the other results, there are clear positive effects on quality even in the partisan election systems, with partisan judges showing the largest increase in caselaw research. The term length effect is actually negative overall in the legislative retention system (Vermont).

the negative direction, from the baseline effect in the full-discretion states. Under mandatory review, judges have little discretion to set their work load so the operation of intrinsic motivation is weaker.

7.4 Effects of Election System Changes

This section reports the effect of changing judicial election systems. Like the term length extension, these reforms weaken extrinsic electoral incentives and give judges more discretion over their time use. As with the intermediate appellate court effects and term length effects, the estimates are generated by estimating Equation (7.2). Tables 6.1 and 6.2 report the results on moving from a non-partisan contested election system to an uncontested election system, while Tables 7.1 and 7.2 report the results on moving from a partisan contested election system to an uncontested election system. As we will see, partisanship matters.

[TABLE 6.1 HERE]

As shown in Table 6.1 Column 3, the weakening of electoral incentives associated with moving from non-partisan to uncontested elections is associated with an increase in performance on various measures. While the number of majority opinions doesn't change, the number of dissenting opinions does. This means judges are voluntarily taking on more work to express their legal and policy views. As with the term length increase, caselaw research increases. There are strong positive effects on the quality of opinions written, as reflected in positive cites, distinguishing cites, discuss cites, and quoted cites. This suggests that judges are working harder on opinions they care about.[20] Concordantly, the tenured judges are overruled less often by later courts, perhaps because their opinions are more persuasive.

As with the IAC and term length changes, intriguingly, there are higher negative cites per opinion. Moreover, the judges are superseded by statutes more often. These results could be interpreted (along with the increased number of dissents), as greater judicial independence (as in Choi et al. 2010). Judges now face less pressure to pander to voters and the political system generally; they decide cases following their own preferences and therefore with more

[20] Another possible interpretation of the citation results is that the composition of cases changes after the reform due to political reasons. Elected judges may not want to take cases that are politically controversial, for example cases about abortion rights. After the reform they take these types of cases and therefore are cited more often. Given the structure of our data and in particular the absence of a good measure for the controversy of a case, we can't satisfactorily test for this possibility. However, we note that the other reforms – the IAC, term length change, and salary change – do not change the politicization of the supreme court yet have similar effects on citations. If our results for the election system changes were due to case composition, we wouldn't expect to have effects for the non-political reforms.

variance. In consequence their reasoning will be criticized by future judges and superseded by voter-driven legislators more often.

[TABLE 6.2 HERE]

Following in the stride of previous sections, Table 6.2 looks at the move from non-partisan to uncontested elections interacted with the level of appellate review discretion in each of these states. Looking first to the Partial Discretion column, it seems that the treatment operated differently on the judges in this state (Maryland). Instead of increasing research and opinion quality, judges in this state decreased the number of opinions and increased opinion length. Meanwhile, there don't seem to be many differences between the mandatory-review states and the full-discretion states from this treatment, which could mean that the judges in this set of mandatory-review states have other means besides formal case selection to control the composition of the case portfolio.

[TABLE 7.1 HERE]

Next we look at the effect of moving from a partisan system to an uncontested system. Before the reform, partisan politicians selected by the political party apparatus ran for election to be state supreme court judges. After the reform, they had de facto tenure; both political parties and voters had far less influence on judge behavior. As shown in Table 7.1, moving from the partisan system to the uncontested system has no positive effects on performance – it actually has some negative effects. In particular, there is a statistically significantly negative within-judge effect on positive cites per opinion, (marginally) distinguishing cites per opinion, discuss cites per opinion, and out-of-state cites per opinion. In response to weakened electoral incentives, performance among these partisan judges actually falls.

[TABLE 7.2 HERE]

Table 7.2 shows that case review discretion also matters for this reform. The decreases in quality after the reform are weaker under partial discretion, proportionally if not statistically. At least by the size of the coefficient, they might even go in the other direction.

There are two main approaches for reconciling the different effects of the non-partisan-to-uncontested and partisan-to-uncontested reforms. First, we can look to the differences in electoral incentives imposed by the non-partisan and partisan systems. As shown in Lim and Snyder (2013), non-partisan contested systems have the most competitive election challenges

where voters and bar associations pay attention to judge campaigning. In the partisan system, people vote based on party affiliation rather than judge campaign activities. In the uncontested system, judges have tenure so there are no campaigning incentives to speak of. In the non-partisan system, therefore, campaigning and other election-related activities are more beneficial and judges allocate time away from judging and toward those activities. We would observe this in the data by the increase in judge performance on moving away from the non-partisan system, with a zero effect on moving away from the partisan system.

However, we don't see a zero after the partisan reform – we see a negative, which can't be explained just by Lim et al.'s hypothesis. Another hypothesis is that judges selected by the non-partisan system are different from those selected by a partisan system. On this view, partisan-selected judges have lower intrinsic preferences for judging. Non-partisan elections, influenced as they are by bar associations and news editorials, select for more technocratic judges that do have an intrinsic motivation to increase decision quality when they have more free time. Thus, when electoral incentives are weakened, non-partisan-selected judges spend more time on their opinions. Partisan-selected judges under mandatory review are busy working through the mandatory caseload so there is no effect from tenure. But under fully discretionary review, opinion quality goes down for partisan-selected judges due to low intrinsic preference for judging.

7.5 Electoral Cycle Effects

To further study the effects of electoral incentives on judge behavior, we look at how judges change their behavior over time in response to the election cycle. In particular, we can exploit the staggered election cycle and compare judges sitting on the same court. The election schedule is arbitrarily assigned by history, so one can reasonably assume that it is uncorrelated with other institutional or socioeconomic factors that would affect individual judge performance. Judges who are up for election have less free time than judges who are not up for election, so if judges have intrinsic motivation to do their job then they will spend less time on it when up for election. Moreover, this effect should be weakest in the uncontested electoral system where judges have de facto tenure.

Formally, we estimate

$$y_{ist} = \text{JUDGE}_i + \text{STATE}_s \times \text{TIME}_t + Z'_{ist}\rho + \epsilon_{ist}$$

where JUDGE_i is a judge fixed effect,and $\text{STATE}_s \times \text{TIME}_t$ is a state-year fixed effect for

each s and year t. The term $Z_{ist} = (Z_{ist}^{NP}, Z_{ist}^{P}, Z_{ist}^{U})$ is a vector of indicators for each of the three electoral systems (non-partisan, partisan, and uncontested) that equals one when judge i is up for election at year t in that system. This regression compares the performance of judges who are up for election to other judges on the same court that are not up for election, controlling for judge-specific characteristics and arbitrary state-level trends. As noted in Section 4, we restrict this analysis to states with random or rotating case assignment (although this doesn't matter for the results). In these states, there is less scope for giving easy cases to the election-cohort judges, so we can interpret changes in output as changes in effort rather than changes in the types of cases a judge hears.

[TABLE 8.1 HERE]

Table 8.1 reports the results on the performance effect of being up for election, with three triads of columns representing the fixed-effects specifications. The first set of columns includes state fixed effects and state trends; the second set of columns includes judge fixed effects and state trends; the third set of columns includes judge fixed effects and state-year fixed effects. Within specification, the abbreviations NP, P, and U stand for non-partisan elections, partisan elections, and uncontested elections, respectively. While the third set of columns is preferred for measuring the causal effect on judges of being up for election, the first set of columns is useful because it summarizes the average effect on the whole court when more judges are up for election in a particular year, relative to trend (although these coefficients also include other election-year factors unrelated to judicial elections).

First consider the effects in NP3, which give the effect of being up for non-partisan election relative to other judges on the court. With less time for judging, how do they reorganize their time? They write (marginally) fewer opinions, significantly fewer dissents, and significantly fewer total words. They are directly reducing the amount of work done. However, this change in work level does not affect opinion length, research, or opinion quality; it seems that they reduce the number of majority and discretionary opinions in order to maintain the quality of their majority opinions. This could be interpreted as evidence that these non-partisan-selected judges care about the quality of their opinions. It is also important to note that these judges are making decisions as a group; judges who are not up for election are likely picking up some of the slack for their colleagues who are up for election.

Next, examine column P3. As with the non-partisan judges, partisan judges when up for election reduce the number of opinions written and number of words written. The number of dissents is not affected, however. Moreover, the length of opinions and caselaw research

decrease during elections, suggesting that part of the time difference comes out of the work put into individual opinions. In turn, there are negative measures on some of the per-opinion citation measures, notably distinguishing cites, discussion cites, quoted cites and out-of-state cites. When up for election, the partisan judges compromise on quality – unlike their non-partisan counterparts. This is consistent with the discussion in Section 7.4 about differences in how these systems select judges.

Finally, consider column U3 on uncontested elections. In contrast to the previous two columns, we don't see any statistically significant negatives. If anything, these judges work a bit harder during their election years, as indicated by an increase in caselaw research and out-of-state citations. These results reflect that with uncontested retention elections, the judges have de facto tenure and don't need to reduce their time allocation to judging based on the electoral cycle.

[TABLE 8.2 HERE]

As a final empirical analysis, we examine differences in the electoral effect due to mandatory case review. Table 8.2 reports these results. There are three sets of columns, but they are divided up by system and each coefficient is from a regression that includes judge fixed effects and state-year fixed effects. As with previous sections, we interact the electoral effect with the appellate-review rule.

In the non-partisan system, the negative effects on majority and dissenting opinions vanish under reduced discretion – these judges don't have as much control over their case portfolios. Correspondingly, the effect on total words written is weaker. With partial discretion, elections have a more negative effect on opinion length and research, perhaps reflecting the effect of less time to work on the mandatory-review cases.

In the partisan system, under partial discretion the negative effect on number of majority opinions is weakened to zero. Instead, these states have a strong negative effect on concurrences written, which could be another margin of reducing work that the full-discretion judges accomplish by reducing majority opinions written. Aside from that, there are mostly zeros in the Partial Discretion column, meaning that the judges in these states reduce work as much as their full-discretion counterparts.

Finally, in the uncontested system, examining the review-standard coefficients shows that case selection matters even under de facto tenure. Under mandatory review, the judges who are up for election write fewer opinions, suggesting that the judges who are not up for election cover their colleague's mandatory caseload to some degree. This makes sense

because even with mandatory appeal, there is generally discretionary assignment of cases across judges. Like the partisan judges under partial discretion, the tenured judges under mandatory review reduce their concurrences in response to electoral demands, however weak. The other estimates are somewhat haphazard and don't contribute to a clear picture.

8 Discussion

The goal of this paper has been to measure the causal effect of changes in employment conditions on judicial behavior. Given that judges have low powered incentives that do not explicitly link pay to performance, a standard agency theory would predict that reducing incentive pressure would have no effect on behavior. We can reject that hypothesis; the reduction of time pressure is associated with better-researched opinions that are cited more often by later judges. When time pressure increases, judges prefer to reduce the number of opinions written rather than compromise on quality. These findings are consistent with the view that judges are professionals who care about the quality of their work, and that at the margin they prefer to maintain high quality at the cost of lower quantity. We have shown not only that judges have intrinsic values, but also that those values include quality appellate opinion writing.

We have also shown that the importance of intrinsic motivation depends on the observable characteristics of courts and judges. Discretion over case selection – that is, the level of control a judge has over his work material – contributes to stronger effects of intrinsic motivation, consistent with the early work of Deci (1971). In addition, the politics of judge appointment seem to matter a lot, with partisan judges responding less well to stronger tenure than their non-partisan counterparts. This suggests that non-partisan systems, where expert bar association evaluations are more influential in the electoral process, select for an intrinsic quality preference more powerfully than partisan systems, where political party affiliations matter most.

These results provide field evidence for theory work on intrinsic motivation (Benabou and Tirole, 2003; Prendergast, 2008; Rebitzer and Taylor, 2011). State supreme court judges are a highly selected group of individuals, concluding a significant career in the law by serving in one of the most prestigious legal positions in their state. Our results support the unsurprising claim that state supreme court judges are motivated by professionalism (White, 1959; Wilensky, 1964), by career concerns (Dewatripont et al., 1999b; Francois, 2000; Prendergast, 2007), or both. Extrinsic incentives that impose time pressure reduce

judge performance by crowding out intrinsic motivation to do a good job, consistent with previous empirical work in other settings (Gneezy et al., 2011).

Researchers in law and economics have begun to recognize that what judges care about matters for their output. For example, Epstein et al. (2013) quote correspondence from economist Andrei Shleifer:

> Consider common law judges who face few prospects of promotion... and cannot be fired or voted out. These judges face almost no incentives. They need to move cases through, and they need to be not so utterly random that they get overturned very much. But these are not enormously strict constraints. So what consequences follow? I think that in this context just about any external or internal motivation can prove decisive.

We have provided empirical evidence of the interaction between such external and internal motivations. By demonstrating that judges care about their opinion output, these findings strengthen the previous empirical studies using features of judicial opinions to measure judicial performance (e.g., Choi et al., 2010; Epstein et al., 2013). Our results support the view that both institutional rules and judge preferences are important inputs into a well-functioning legal system, which might partly explain the significant cross-country variation in the quality of legal systems (Djankov et al., 2003).

The recent work in behavioral economics has demonstrated a great deal of variation in individual preferences. In particular, this work has shown that the standard agency model's assumption that agents are motivated only by pecuniary returns is a (useful) simplification.[21] It remains an open question the extent to which we could design institutions that select individuals with particular social preferences, and in particular a preference to act in the public interest.[22] Our results suggest that state appellate court judges are one group of officials where institutions do matter for selection of preferences.

References

Alesina, A. and G. Tabellini (2007). Bureaucrats or politicians? part i: A single policy task. *The American Economic Review 97*(1), 169–179.

[21] See Fehr and Schmidt (1999) for a synthetic discussion of social preferences.

[22] See Lazear (1989) for an interesting paper that discusses how firms may choose employees with particular preferences. Whether workers are intrinsically motived also matters for the design of labor market policy (Babcock et al., 2012).

Alesina, A. and G. Tabellini (2008). Bureaucrats or politicians? part ii: Multiple policy tasks. *Journal of Public Economics 92*(3-4), 426–447.

Angrist, J. and J.-S. Pischke (2009). *Mainly Harmless Econometrics*. Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010, Spring). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Autor, D. H., J. J. Donohue, and S. J. Schwab (2004, May). The employment consequences of wrongful-discharge laws: Large, small, or none at all? *American Economic Review 94*(2), 440–446.

Autor, D. H., J. J. Donohue, and S. J. Schwab (2006, May). The costs of wrongful-discharge laws. *Review of Economics and Statistcs 88*(2), 211–231.

Avraham, R., L. S. Dafny, and M. M. Schanzenbach (2013). The impact of tort reform on employer-sponsored health insurance premiums. *Journal of Law Economics & Organization 29*.

Babcock, L., W. J. Congdon, L. F. Katz, and S. Mullainathan (2012). Notes on behavioral economics and labor market policy. *IZA Journal of Labor Policy 1*.

Baker, S. and C. Mezzetti (2012). A theory of rational jurisprudence. *Journal of Political Economy 120*(3), pp. 513–551.

Baker, S. A. (2008). Should we pay federal circuit judges more? *BUL Rev. 88*, 63.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2014). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. Technical report.

Benabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3), 871–915.

Benabou, R. and J. Tirole (2003, JUL). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Benabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bertrand, M., E. Duflo, and S. Mullainathan (2004, Feb). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Besley, T. and S. Coate (2003). Elected versus appointed regulators: Theory and evidence. *Journal of the European Economic Association 1*(5), 1176–1206.

Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *The*

American economic review 95(3), 616–636.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2).

Carvell, D., J. Currie, and W. B. MacLeod (2012, Spring). Accidental death and the rule of joint and several liability. *Rand Journal of Economics 43*(1), 51–77.

Chandra, A., D. Cutler, and Z. Song (2012). Chapter six - who ordered that? the economics of treatment choices in medical care. In T. G. M. Mark V. Pauly and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2, pp. 397 – 432. Elsevier.

Choi, S., M. Gulati, and E. Posner (2008). Judicial evaluations and information forcing: Ranking state high courts and their judges. *Duke LJ 58*, 1313.

Choi, S. J., G. M. Gulati, and E. A. Posner (2009). Are judges overpaid-a skeptical response to the judicial salary debate. *Journal of Legal Analysis 1*, 47.

Choi, S. J., G. M. Gulati, and E. A. Posner (2010). Professionals or politicians: The uncertain empirical case for an elected rather than appointed judiciary. *Journal of Law, Economics, and Organization 26*(2), 290.

Christensen, R. K., J. Szmer, and J. M. Stritch (2012). Race and gender bias in three administrative contexts: Impact on work assignments in state supreme courts. *Journal of Public Administration Research and Theory*.

Coviello, D., A. Ichino, and N. Persico (2011). The inefficiency of worker time use. Technical report, NYU.

Currie, J. and W. B. MacLeod (2008, May). First do no harm? tort reform and birth outcomes. *Quarterly Journal of Economics 123*(2), 795–830.

Currie, J. M. and W. B. MacLeod (2013, April). Diagnosis and unnecessary procedure use: Evidence from c-section. Technical Report 18977, NBER, Cambridge, MA.

Dal Bo, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *Quarterly Journal of Economics 128*(3), 1169–1218.

Dal Bó, E. and M. Rossi (2011). Term length and the effort of politicians. *The Review of Economic Studies 78*(4), 1237–1263.

D'Aspremont, C. and L.-A. Gerard-Varet (1979, February). Incentives and incomplete information. *Journal of Public Economics 11*(1), 25–45.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic montivation. *Journal of Personality and Social Psychology 18*(1), 105–115.

Delfgaauw, J. and R. Dur (2008). Incentives and workers' motivation in the public sector.

The Economic Journal 118(525), pp. 171–191.

Dewatripont, M., I. Jewitt, and J. Tirole (1999a). The economics of career concerns, part i: Comparing information structures. *REStudies 66*(1), 183–98.

Dewatripont, M., I. Jewitt, and J. Tirole (1999b). The economics of career concerns, part ii: Application to missions and accountability of government agencies. *The Review of Economic Studies 66*(1), pp. 199–217.

Djankov, S., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2003). Courts. *Quarterly Journal of Economics 118*(2), 453–517.

Eisenberg, T. and G. P. Miller (2009). Reversal, dissent, and variability in state supreme courts: The centrality of jurisdictional source. *Boston University Law Review 89*, 1451–1504.

Epstein, L. and J. Knight (2013). Reconsidering judicial preferences. *Annual Review of Political Science 16*(1), 11–31.

Epstein, L., W. M. Landes, and R. A. Posner (2013). *The Behavior of Federal Judges*. Harvard University Press.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*(3), 817–68.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice 50*(1-3), 5–25.

Francois, P. (2000). Public service motivation as an argument for government provision. *Journal of Public Economics 78*(3), 275–299.

Francois, P. and M. Vlassopoulos (2008, February). Pro-social motivation and the delivery of social services. *CESifo Economic Studies 54*(1), 22–56.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review 87*(4), 746–755.

Gennaioli, N. and A. Shleifer (2007). The evolution of common law. *Journal of Political Economy 115*(1), 43–68.

Glaeser, E. L., S. Johnson, and A. Shleifer (2001). Coase versus the coasians. *Quarterly Journal of Economics 116*(3), 853–99.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives 25*(4), 191–209.

Gneezy, U. and A. Rustichini (2000, Aug). Pay enough or don't pay at all. *Quarterly Journal of Economics 115*(3), 791–810.

Gordon, S. and G. Huber (2007). The effect of electoral competitiveness on incumbent

behavior. *Quarterly Journal of Political Science 2*(2), 107–138.

Hall, M. and C. Bonneau (2006, January). Does quality matter? challengers in state supreme court elections. *American Journal of Political Science 50*(1), 20–33.

Hanssen, F. (2004). Is there a politically optimal level of judicial independence? *The American Economic Review 94*(3), 712–729.

Iaryczower, M., G. Lewis, and M. Shum (2013). To elect or to appoint? bias, information, and responsiveness of bureaucrats and politicians. *Journal of Public Economics 97*, 230–244.

Kerr, S. (1975, December). On the folly of rewarding A, while hoping for B. *Academy of Management Journal 18*(4), 769–783.

Klerman, D. and P. Mahoney (2005). The value of judicial independence: Evidence from eighteenth century england. *American Law and Economics Review 7*(1), 1.

Knight, J. and L. Epstein (1996, NOV). The norm of stare decisis. *American Journal oOf Political Science 40*(4), 1018–1035.

Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *The American Economic Review 87*(2), pp. 359–364.

Lacetera, N. and M. Macis (2010). Do all material incentives for pro-social activities backfire? the response to cash and non-cash incentives for blood donations. *Journal of Economic Psychology 31*(4), 738 – 748.

Landes, W. M. and R. A. Posner (1980). Legal change, judicial behavior, and the diversity jurisdiction. *The Journal of Legal Studies 9*(2), pp. 367–386.

Lazear, E. P. (1989, December). Pay equality and industrial politics. *Journal of Political Economy 97*, 561–580.

Lim, C. and J. Snyder (2013). Elections and government accountability: Evidence from the us state courts. mimeo, Havard University.

Lim, C. H. S. (2013). Preferences and incentives of appointed and elected public officials: Evidence from state trial court judges. *American Economics Review*.

MacLeod, W. B. and V. Nakavachara (2007, June). Legal default rules: the case of wrongful discharge laws. *Economic Journal 117*, F1–F62.

Mas, A. (2006, AUG). Pay, reference points, and police performance. *Quarterly Journal of Economics 121*(3), 783–821.

Miles, T. J. (2000, Apr). Common law exceptions to employment at will and us labor markets. *Journal of Law Economics & Organization 16*(1), 74–101.

Milkovich, G. T. and A. K. Wigdor (1991). *Pay for Performance: Evaluating Performance*

and Appraisal Merit Pay. Washington, D.C., U.S.A.: National Academy Press.

Prendergast, C. (2007, March). The motivation and bias of bureaucrats. *American Economic Review 97*(1), 180–196.

Prendergast, C. (2008, May). Intrinsic motivation and incentives. *The American Economic Review 98*(2), 201–205.

Rebitzer, J. B. and L. J. Taylor (2011). Extrinsic rewards and intrinsic motives: Standard and behavioral approaches to agency and labor markets. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Volume 4*, Volume 4, Chapter 8. Elsevier.

Shepherd, J. (2009a). Are appointed judges strategic too? *Duke Law Journal 58*(7), 1589–1626.

Shepherd, J. (2009b). The influence of retention politics on judges' voting. *The Journal of Legal Studies 38*(1), 169–206.

Spottswood, M. (2007). Free speech and due process problems in the regulation and financing of judicial election campaigns. *Nw. UL Rev. 101*, 331.

Tabarrok, A. and E. Helland (1999). Court politics: The political economy of tort awards. *Journal of Law and Economics 42*(1), pp. 157–188.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*.

Wilensky, H. L. (1964). The professionalization of everyone? *American Journal of Sociology 70*(2), pp. 137–158.

A Theory Appendix: Solution to Judge Optimization Problem

A.1 General Case

We begin by considering a general time allocation problem, and then we apply the solution to the judge's problem. There are n activities, indexed by $i \in \{1, .., n\}$. The time allocated to each activity is given by T_i, with vector representation $\vec{T} = \{T_1, ..., T_n\}$. Let $\bar{T} > 0$ be the total time available. The gain from each activity is represented by $U_i(T_i, \alpha_i)$, with the

corresponding payoff function given by:

$$u\left(\vec{T},\vec{\alpha}\right) = \sum_{i=1}^{n}\beta_i \log\left(U_i\left(T_i,\alpha_i\right)\right), \tag{A.1}$$

$$= \sum_{i=1}^{n}\beta_i u_i\left(T_i,\alpha_i\right). \tag{A.2}$$

Note that since U_i is (weakly) concave in T_i, then u_i is strictly concave in T_i (a strictly concave function of a concave function is strictly concave). The sum of concave functions is concave, so $u(\cdot)$ is concave in $\vec{T}$. The agent faces the following time constraints:

$$\sum_{i=1}^{n} T_i \leq \bar{T} \tag{A.3}$$

$$T_i \geq 0, \forall i. \tag{A.4}$$

The objective is to allocate time across activities to maximize the payoff function subject to these constraints.

Existence and uniqueness follow from the standard concave optimization assumptions over a compact, convex set. We can characterize the optimum using the first order conditions for the Lagrangian:

$$L = u\left(\vec{T},\vec{\alpha}\right) + \mu\left(\bar{T} - \sum_{i=1}^{n} T_i\right).$$

Let $\vec{T}^*\left(\vec{\alpha}\right)$ denote the optimum, and let $\mu^*\left(\vec{\alpha}\right)$ be the associated Lagrange multiplier. Here we ignore the non-negativity constraints $T_i \geq 0$ and consider the conditions for an interior solution. The first-order condition for T_i is given by:

$$0 = L_{T_i} = \beta_i \frac{\partial u_i}{\partial T_i} - \mu^*,$$

$$\beta_i \frac{\partial u_i}{\partial T_i} = \mu^*, \tag{A.5}$$

and hence we have at the optimum:

$$\frac{\partial u_i/\partial T_i}{\partial u_j/\partial T_j} = \frac{\beta_j}{\beta_i}, \forall i \neq j. \tag{A.6}$$

Next we work out the comparative static conditions. Let $J = \nabla u$ be the Jacobian for u and H be the corresponding Hessian matrix ($H_{ij} = \frac{\partial^2 u}{\partial T_i \partial T_j}$). The first order conditions can

be written in matrix form:

$$\left[\begin{array}{c} J - \mu^* \vec{1} \\ \bar{T} - \vec{1}^T \vec{T}^* \end{array} \right] =,$$

where $\vec{1}$ is a vector of ones and T denotes the transpose. We take derivatives with respect to α_1 (the other cases are similar – this is easier to write), which gives us:

$$\left[\begin{array}{cc} H & -\vec{1} \\ -\vec{1}^T & 0 \end{array} \right] \left[\begin{array}{c} \frac{\partial \vec{T}}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{array} \right] + \left[\begin{array}{c} \beta_1 \frac{\partial^2 u_1}{\partial T_1 \partial \alpha_1} \\ 0 \\ \vdots \\ 0 \end{array} \right] = \vec{0}. \tag{A.7}$$

Multiply each side by the vector of effects:

$$\left[\begin{array}{c} \frac{\partial \vec{T}}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{array} \right]^T \left[\begin{array}{cc} H & -\vec{1} \\ -\vec{1}^T & 0 \end{array} \right] \left[\begin{array}{c} \frac{\partial \vec{T}}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{array} \right] + \left[\begin{array}{c} \frac{\partial \vec{T}}{\partial \alpha_1} \\ \frac{\partial \mu^*}{\partial \alpha_1} \end{array} \right]^T \left[\begin{array}{c} -\beta \frac{\partial^2 u_1}{\partial T_1 \partial \alpha_1} \\ 0 \\ \vdots \\ 0 \end{array} \right] = \vec{0}. \tag{A.8}$$

Since the time constraint is always binding we have $\sum_{i=1}^n \frac{\partial T_i}{\partial \alpha_i} = \frac{\partial \bar{T}}{\partial \alpha_i} = 0$ and (A.8) implies:

$$\frac{\partial \vec{T}}{\partial \alpha_1}^T H \frac{\partial \vec{T}}{\partial \alpha_1} = -\beta_1 \frac{\partial^2 u_1}{\partial T_1 \partial \alpha_1} \times \frac{\partial T_1}{\partial \alpha_1}. \tag{A.9}$$

The first term on the right-hand side is negative since the objective function is strictly concave, and hence H is negative definite. Thus for a general activity i we conclude:

$$\begin{aligned} sign(\frac{\partial T_i^*}{\partial \alpha_i}) &= -sign\left(\frac{\partial \vec{T}}{\partial \alpha_1}^T H \frac{\partial \vec{T}}{\partial \alpha_1} \right) \times sign\left(\beta_1 \frac{\partial^2 u_1}{\partial T_1 \partial \alpha_1} \right) \\ &= sign\left(\beta_1 \frac{\partial^2 u_1}{\partial T_1 \partial \alpha_1} \right) > 0. \end{aligned}$$

The final line follows from the assumption that the parameters have a positive effect. Thus we have $\frac{\partial T_i^*}{\partial \alpha_i} > 0$. The binding time constraint implies that for some $j \neq i$ we have $\frac{\partial T_j^*}{\partial \alpha_i} < 0$. From (A.5) this implies $\frac{\partial \mu^*}{\partial \alpha_i} > 0$, and which combined with the strict concavity of u_j in T_j implies that $\frac{\partial T_j^*}{\partial \alpha_i} < 0$ for all $j \neq i$ such that $T_j^* > 0$. This completes the proof of Proposition 4.

A.2 Time Allocation to Cases Sub-problem

The judging sub-problem is the solution to:

$$U_J\left(T_J, \alpha_J\right) = \max_{\gamma, \delta(.)} V_J\left(\gamma, \delta\left(\cdot\right), \alpha_J\right) \tag{A.10}$$

subject to:

$$\alpha_J \int_{\gamma}^{1} \left(t_0 + \delta\left(\gamma\right)\right) f\left(\gamma\right) d\gamma + \left(1 - \alpha_J\right) \int_{0}^{1} \left(\left(V_0 + \lambda V_1\left(\delta\left(\gamma\right), \gamma\right)\right)\right) \quad \leq \quad T_J. \tag{A.11}$$

Let μ_J be the multiplier for the time constraint. The Lagrangian can be written as:

$$\begin{aligned} L \quad = \quad & \int_{0}^{1} \left(\left(V_0 + \lambda V_1\left(\delta\left(\gamma\right), \gamma\right)\right)\right) - \mu_J \left(t_0 + \delta\left(\gamma\right)\right) f\left(\gamma\right) d\gamma \\ & -\alpha_j \int_{0}^{\gamma} \left(\left(V_0 + \lambda V_1\left(\delta\left(\gamma\right), \gamma\right)\right)\right) - \mu_J \left(t_0 + \delta\left(\gamma\right)\right) f\left(\gamma\right) d\gamma \\ & + \int_{\gamma}^{1} \left(\left(V_0 + \lambda V_1\left(\delta\left(\gamma\right), \gamma\right)\right)\right) - \mu_J \left(t_0 + \delta\left(\gamma\right)\right) f\left(\gamma\right) d\gamma \\ & + \mu_J T_J \end{aligned} \tag{A.12}$$

The first observation is that

$$\frac{\partial U_J\left(T_J\right)}{\partial T_J} = \mu_J. \tag{A.13}$$

Namely, the multiplier gives us the marginal utility of time allocated to judging. As time becomes more expensive, then this multiplier is larger, which will have some clear predictions regarding the allocation of time to cases. The same multiplier appears in each subcase, and so we can take it as a free parameter that is adjusted so that the total budget constraint is satisfied. We can write $T_i\left(\mu_J\right)$, for $i \in \{L, J, A\}$, which are increasing functions of μ_J.

Next, the first order condition for the allocation of time to a case is given by:

$$\frac{\partial V_1\left(\delta^*\left(\gamma, \mu_J\right), \gamma\right)}{\partial \delta} = \frac{\mu_J}{\lambda}. \tag{A.14}$$

Notice that the time allocated per case does not vary directly with α_J, only with the value of time (which in turn is affected by α_J). When $\lambda \to 0$, the excess time allocated to a case goes to zero and the effort per case is insensitive to the cost of time. Moreover, we have for $\lambda > 0$:

$$\frac{\partial \delta^*\left(\gamma, \mu_J\right)}{\partial \mu_J} < 0.$$

The time per case falls as time is more constrained.

The first order condition for the number of cases ($\curlywedge$) is given by:

$$\frac{V_0 + \lambda V_1\left(\delta^*\left(\curlywedge^*, \mu_J\right), \curlywedge^*\right)}{t_0 + \delta^*\left(\curlywedge^*, \mu_J\right)} = \mu_J. \tag{A.15}$$

This condition is independent of α_J. It says that, at the optimum, the judge is indifferent between reviewing and not reviewing the marginal case of importance $\curlywedge$. Applying the envelope theorem and A.14 we get:

$$\frac{\partial \curlywedge^*}{\partial \mu_J} = \frac{t_0 + \delta^*}{\lambda(\partial V_1/\partial \gamma)} > 0.$$

As time becomes more scarce, the judge hears fewer cases (chooses a higher threshold $\curlywedge$). Also observe that, holding the price of time μ_J fixed, increasing λ results in the judge hearing more cases, as well as spending more time per case.

Finally, since the time constraint is binding, then the utility is equal to the Lagrangian:

$$U_J\left(T_j, \alpha_J\right) = L\left(\delta^*\left(\gamma, \mu_J\left(T_J\right)\right), \curlywedge^*, \mu_J\left(T_J\right)\right),$$

where $\frac{\partial \mu_J(T_J)}{\partial T_J} < 0$,and μ_J is given by (A.13). Then we have the total number of cases as a function of μ_J:

$$N^*\left(\mu_J, \alpha_J\right) = \int_0^1 f\left(\gamma\right) d\gamma - \alpha_J \int_0^{\curlywedge^*(\mu_J)} f\left(\gamma\right) d\gamma, \tag{A.16}$$

which implies:

$$\frac{\partial N^*}{\partial \mu_J}, \quad \frac{\partial N^*}{\partial \alpha_J} < \quad 0.$$

With stricter time constraints or more discretion over case selection, the number of cases falls.

The total time spent on judging is given by:

$$T_J\left(\mu_J, \alpha_J\right) = \int_0^1 \left(t_0 + \delta^*\left(\gamma, \mu_J\right)\right) f\left(\gamma\right) d\gamma - \alpha_J \int_0^{\curlywedge^*(\mu_J)} \left(t_0 + \delta^*\left(\gamma, \mu_J\right)\right) f\left(\gamma\right) d\gamma. \tag{A.17}$$

An increase in α_J corresponds to an increase in discretion, and a decrease in time use holding the cost of time fixed. Thus we have:

$$\frac{\partial T_J}{\partial \alpha_J}, \frac{\partial T_J}{\partial \mu_J} < 0.$$

We also have the budget constraint:

$$T_J\left(\mu_J, \alpha_J\right) = T_J.$$

This allows us to write $\mu_J\left(\alpha_J\right)$, from which we get $\frac{\partial \mu_J}{\partial \alpha_J} < 0$. Hence with (A.13), this implies:

$$\frac{\partial^2 U_J}{\partial T_J \partial \alpha_J} > 0.$$

When V_0 is sufficiently large this implies:

$$\frac{\partial^2 u_J}{\partial T_J \partial \alpha_J} > 0.$$

This is a theoretical explanation for Deci's (1971) empirical results on the motivating effect of reducing contingency.

The increase in discretion increases the time spent per case, and since it decreases the cost of time, then the time per case goes up. The effect on the number of cases is:

$$dN^*\left(\mu_J\left(\alpha_J\right), \alpha_J\right)/d\alpha_J = \frac{\partial N^*}{\partial \mu_J}\frac{d\mu_J}{d\alpha_J} + \frac{\partial N^*}{\partial \alpha_J} < 0.$$

There are fewer low quality cases, but more high-quality cases that consume more time. Finally, when the intrinsic value of judging is low, then there is little variation in time spent on cases, and variations in the cost of time are reflected mainly in the case load.

TABLE 1.1
State Supreme Court Employment Conditions Changes

Rule Change	Description	Successful Policy Changes
Intermediate Appellate Court	Before, state supreme court judges reviewed a case directly from trial, with mandatory review. After, an intermediate court reviewed the case first, and the court exercised discretionary review.	FL (1956), MI (1963), AZ (1964), NM (1965), MD (1966), NC (1967), OK (1967), AL (1969), OR (1969), WA (1969), CO (1970), MA (1972), KY (1975), IA (1976), KS (1976), WI (1977), AR (1978), HI (1979), AK (1980), ID (1981), CT (1982), MN (1983), VA (1984), ND (1987), UT (1987), NE (1990)
Compensation Increase	Increases in judge annual salary due to legislation.	All states
Term Length Increase	Increase in the length of term of office before facing re-election.	IL (1962), HI (1968), IN (1970), MT (1972), SD (1972), VT (1974), KY(1975)
Term Length Decrease	Decrease in the length of term of office before facing re-election.	PA (1968), LA (1974), MD (1976)
Non-partisan contested to uncontested re-election	Before, judges face a challenger, but party affiliations not on ballot. After, judges face no challenger.	AZ (1974), WY (1972), FL (1976), MD (1976), SD (1980), UT (1985)
Partisan contested to uncontested re-election	Before, judges face a challenger with party affiliations on ballot. After, judges face no challenger.	KA (1958), IA (1962), NE (1962), IL (1964), IN (1970), CO (1966), OK (1967), TN (1971), NM (1988)

Notes. This table summarizes the changes in state supreme court employment conditions that are used as treatments in our empirical analysis. Column 1 names the rule change, Column 2 gives a brief description, and Column 3 lists the states that enacted the rule change and the year that they did so. The following reforms are not the focus of the analysis but are included as control variables: establishment of a court administrative office, moving from partisan contested election to nonpartisan contested election, moving from partisan contested election to governor retention, increase in the number of judges, moving from single-terms to uncontested election, and establishment of a mandatory retirement age.

TABLE 1.2
State Supreme Court Rules of Appellate Procedure

Procedural Rule	Description	List of States
Case Selection		
Full Discretion	State supreme court judges have full discretion in whether to review most non-death-penalty cases.	CA, CT, FL, HI, IA, ID, IL, IN, LA, MI, MT, NC, NH, NY, OK, TN, TX, VA, WA, WI, WV
Partial Discretion	State supreme court judges have discretionary review for some types of cases, but a significant proportion of cases require review.	AK, AL, AR, CO, DE, GA, KS, KY, MA, MD, MN, MO, NE, NJ, NM, OR, PA, RI, SC
Mandatory Review	State supreme court judges have no discretion whether to accept for review; some review is always mandatory.	AZ, ME, MS, ND, NV, SD, UT, VT, WY
Case Assignment		
Random	Cases are randomly assigned to authoring judges.	ID, LA, MI, MS, NH, NY, OH, SD, TN, TX, VA, WA, WI
Rotation	Cases are assigned to judges on a rotating basis.	AL, AK, AR, FL, GA, IL, IA, ME, MN, MO, MT, NE, NV, NM, NC, ND, OK, RI, SC, UT, VT, WV
Discretion	The chief justice chooses which judge receives each case.	AZ, CA, CO, CT, DE, HI, IN, KS, KY, MD, MA, NJ, OR, PA, WY

Notes. This table summarizes two dimensions of state court appellate procedure that play a role in our empirical analysis. Column 1 names the rule, Column 2 describes it, and Column 3 lists the states that follow that rule in their supreme court.

TABLE 2.1
Summary Statistics on Judge Characteristics

	All Judges		Partisan Election		Non-Partisan Election		Uncontested Election	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Start Age	53.26	8.44	53.31	8.81	53.01	8.20	52.10	7.74
Career Length	12.69	8.07	12.10	8.31	13.07	8.18	12.99	6.82
Female	0.05	0.21	0.03	0.17	0.06	0.24	0.06	0.24
Republican	0.39	0.49	0.30	0.46	0.47	0.50	0.33	0.48
Top School	0.15	0.35	0.10	0.30	0.10	0.30	0.10	0.31
Promoted	0.04	0.19	0.04	0.19	0.05	0.22	0.07	0.26
Retire/Resign	0.73	0.44	0.68	0.47	0.75	0.43	0.75	0.43
Died In Office	0.16	0.36	0.20	0.40	0.14	0.35	0.10	0.30
Lost Election	0.06	0.23	0.07	0.26	0.05	0.22	0.03	0.18
N	1700		675		388		219	

Notes. Biographical information by judge election system. Observation is a judge. "All Judges" column also includes judges selected by governor appointment and by legislative appointment. *Start Age* is judge age upon joining the court. *Career Length* is number of years working on the court, conditional on having left the court before 2014. *Female* is a dummy for being female. *Republican* is a dummy for being Republican, conditional on being Republican or Democrat. We have party affiliation for 599 judges or 35.2% of the sample. The other 64.8% are independent or we could not find information on their party affiliation. *Top School* means the judge attended law school at Yale, Harvard, Columbia, Stanford, or Chicago. *Promoted* means they became a judge on a federal court. *Retire/Resign* means they left the court voluntarily. *Died in Office* means they died while sitting on the court. *Lost Election* means they were removed because they lost election.

TABLE 2.2
Summary Statistics on Institutional Variables

	Mean	Std. Dev.	Min	Max
Intermediate Appellate Court	0.72	0.45	0.00	1.00
Real Salary	621.61	109.96	399.97	1177.73
Number of Judges	6.49	1.27	3.00	9.00
Term Length	8.29	2.60	2.00	21.00
Non-Partisan Elections	0.25	0.43	0.00	1.00
Partisan Elections	0.31	0.46	0.00	1.00
Uncontested Election	0.20	0.40	0.00	1.00
Governor Retention	0.09	0.28	0.00	1.00
Life Tenure	0.08	0.27	0.00	1.00
Legislative Retention	0.06	0.24	0.00	1.00
Min Age Requirement	0.32	0.47	0.00	1.00
Max Age Requirement	0.57	0.49	0.00	1.00
Partial Discretion	0.38	0.48	0.00	1.00
Mandatory Review	0.18	0.39	0.00	1.00
Random Assignment	0.26	0.44	0.00	1.00
Rotating Assignment	0.44	0.50	0.00	1.00

Notes. Observation is a state-year. *Intermediate Appellate Court* is a dummy for state-years where the court had an IAC operating. *Real Salary* is deflated by 1984 dollars. *Number of Judges* is the number of judge positions on the court (rather than the number actually sitting due to vacancies). *Term Length* is the number of years in between elections or reappointment. *Nonpartisan Election*, *Partisan Election*, *Uncontested Election*, *Governor Retention*, *Life Tenure*, and *Legislative Retention* are dummies for the methods of judge retention described in Section 4. *Min Age Requirement* is a dummy equaling one in state-years that judges have to be above a certain age to join the court. *Max Age Requirement* is a dummy for state-years that judges have a mandatory retirement age. *Partial Discretion* is a dummy for having partial discretion in selecting cases for review. *Mandatory Review* is a dummy for having to review all cases. This leaves 44.4% of the state-years left over for discretionary review. *Random Assignment* is a dummy for courts that assign cases to judges randomly, and *Rotating Assignment* is a dummy for courts that assign cases to judges by a set rotation. This leaves 30% of the state-years left over for discretionary case assignment.

TABLE 2.3
Summary Statistics on Case Characteristics

	Full Sample		Authored Majority Opinions	
Outcome Variable	Mean	Std. Dev.	Mean	Std. Dev.
Majority Opinion Length (Words)	984.38	1663.96	2228.50	1894.34
Table of Cases Length	9.01	18.58	20.59	24.04
Positive Cites	5.19	19.35	11.74	27.81
Distinguishing Cites	0.77	3.21	1.82	4.82
Negative Cites	0.20	1.05	0.47	1.59
Discuss Cites	1.15	3.39	2.64	4.92
Quoted Cites	1.23	4.88	2.83	7.20
Out-of-State Cites	0.76	3.94	1.73	5.94
Superseded by Statute Cites	0.02	0.17	0.04	0.26
Overruling Cites	0.02	0.38	0.04	0.58
Affirm	0.24	0.41	0.50	0.48
Reverse	0.12	0.31	0.25	0.42
Remand	0.10	0.29	0.20	0.39
Modify	0.03	0.16	0.06	0.23
Civil Law Case	0.28	0.45	0.50	0.50
Criminal Law Case	0.22	0.42	0.30	0.46
Administrative Law Case	0.04	0.20	0.08	0.26
Constitutional Law Case	0.05	0.22	0.10	0.29
Number of Discretionary Opinions	0.10	0.37	0.22	0.53
One Dissent	0.05	0.21	0.10	0.30
Two Dissents	0.01	0.07	0.01	0.11
Three Dissents	0.00	0.03	0.00	0.04
One Concurrence	0.02	0.15	0.05	0.22
Two Concurrences	0.00	0.05	0.01	0.07
Three Concurrences	0.00	0.02	0.00	0.03
One Concurring/Dissenting Opinion	0.01	0.09	0.02	0.13
Two Concurring/Dissenting Opinions	0.00	0.03	0.00	0.04
Number of Observations	1025461		387905	

Notes. Means and standard deviations of case data. The first pair of columns are from the full sample of cases, while the second pair of columns are from the set of authored majority opinions, at least 7 sentences long, that are used in the empirical analysis. See accompanying text for definitions of variables.

TABLE 2.4

Summary Correlations of Case Characteristics

	Maj. Op. Length	TOC Length	Pos Cites	Dist Cites	Neg Cites	Discuss Cites	Quoted Cites	Out-State Cites	Supersede Cites	Overrule Cites	Dissents	Concurs
Maj. Op. Length	1.000											
TOC Length	0.744	1.000										
Pos Cites	0.218	0.232	1.000									
Dist Cites	0.329	0.350	0.531	1.000								
Neg Cites	0.276	0.309	0.462	0.717	1.000							
Discuss Cites	0.308	0.314	0.658	0.537	0.412	1.000						
Quoted Cites	0.276	0.293	0.752	0.483	0.401	0.682	1.000					
Out-State Cites	0.156	0.168	0.353	0.277	0.173	0.398	0.447	1.000				
Supersede Cites	0.154	0.176	0.171	0.363	0.328	0.218	0.179	0.098	1.000			
Overrule Cites	0.081	0.096	0.056	0.106	0.227	0.068	0.062	0.033	0.059	1.000		
Dissents	0.092	0.081	0.075	0.114	0.114	0.125	0.080	0.039	0.057	0.043	1.000	
Concurrences	0.094	0.088	0.070	0.092	0.095	0.116	0.078	0.040	0.051	0.030	0.123	1.000

Notes. Observation is a case, N=387,905. Includes majority opinions that are authored by a state supreme court judge and at least seven sentences long. See accompanying text for variable definitions.

TABLE 2.5
Summary Statistics on Judge-Year Performance Variables

Outcome Variable	Mean	Std. Dev.	Min	Max
Number of Majority Opinions Written	24.98	16.11	1.00	330.00
Number of Dissents Written	3.76	5.70	0.00	153.00
Number of Concurrences Written	1.84	3.48	0.00	60.00
Total Words Written	63831.30	40462.83	298.00	429770.00
Average Length of Majority Opinion	2485.19	1355.36	47.00	38240.00
Average Length of Table of Cases	22.95	17.07	0.00	557.50
Positive Cites Per Opinion	13.26	13.14	0.00	373.00
Distinguishing Cites Per Opinion	2.18	2.78	0.00	51.59
Negative Cites Per Opinion	0.59	0.91	0.00	20.63
Discuss Cites Per Opinion	3.01	2.78	0.00	105.50
Quoted Cites Per Opinion	3.36	4.28	0.00	196.50
Out-of-State Cites Per Opinion	1.86	2.54	0.00	82.43
Cases Overruled	1.08	3.22	0.00	116.00
Cases Superseded by Statute	0.97	1.89	0.00	28.00
Total Positive Cites	293.74	277.46	0.00	6134.00
Total Distinguishing Cites	45.63	56.26	0.00	877.00
Total Negative Cites	11.82	16.12	0.00	246.00
Total Discuss Cites	66.04	54.87	0.00	902.00
Total Quoted Cites	70.86	73.27	0.00	1527.00
Total Out-of-State Cites	43.21	77.26	0.00	2308.00
Total Federal Circuit Cites	16.94	107.92	0.00	3503.00

Notes. Observation is a judge-year, N=15,486. These statistics are constructed from each judge's yearly output of cases. "Per Opinion" measures are divided by the number of majority opinions written that year. See variable definitions in the accompanying text.

TABLE 2.6

Judge-Year Output Summary Statistics By Election System

	Partisan Election		Non-Partisan Election		Uncontested Election	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Number of Majority Opinions Written	27.73	18.51	23.57	13.99	25.90	16.52
Number of Dissents Written	4.39	6.76	3.46	4.57	4.67	5.89
Number of Concurrences Written	2.04	3.98	1.78	2.75	2.61	4.25
Total Words Written	61290.92	37925.77	63201.24	40107.19	78377.02	44958.05
Average Length of Majority Opinion	2143.13	999.21	2548.67	1189.56	2897.53	1605.04
Average Length of Table of Cases	18.69	12.54	22.15	11.99	29.56	21.90
Positive Cites Per Opinion	13.09	14.28	10.45	6.26	14.78	9.79
Distinguishing Cites Per Opinion	1.94	2.57	1.51	1.77	3.39	3.83
Negative Cites Per Opinion	0.63	1.08	0.40	0.54	0.81	1.04
Discuss Cites Per Opinion	2.68	2.50	2.80	1.81	3.55	2.25
Quoted Cites Per Opinion	2.96	3.72	2.58	2.05	3.83	3.21
Out-of-State Cites Per Opinion	2.19	3.42	1.46	0.94	1.83	1.86
Cases Overruled	1.28	3.72	1.03	2.94	1.30	3.39
Cases Superseded by Statute	0.79	1.44	0.74	1.33	1.74	3.01
N	5545		3503		3357	

Notes. Observation is a judge-year. These statistics are constructed from each judge's yearly output of cases. "Per Opinion" measures are divided by the number of majority opinions written that year. See variable definitions in the accompanying text.

TABLE 2.7

Summary Correlations by Judge-Year Output Measure

	# Maj Ops	# Dis Ops	# Con Ops	Total Words	Maj Op Length	TOC Length	Pos Cites	Dist Cites	Neg Cites	Discuss Cites	Quoted Cites	Out-State Cites
# Maj Ops	1.000											
# Dis Ops	0.216	1.000										
# Con Ops	0.197	0.622	1.000									
Total Words	0.576	0.435	0.403	1.000								
Maj Op Length	-0.292	-0.017	0.034	0.368	1.000							
TOC Length	-0.213	0.007	0.059	0.340	0.811	1.000						
Pos Cites	-0.177	0.135	0.128	0.126	0.392	0.417	1.000					
Dist Cites	-0.196	0.100	0.085	0.177	0.426	0.464	0.607	1.000				
Neg Cites	-0.195	0.099	0.085	0.116	0.352	0.399	0.567	0.866	1.000			
Discuss Cites	-0.204	0.124	0.157	0.158	0.505	0.511	0.793	0.495	0.426	1.000		
Quoted Cites	-0.189	0.094	0.099	0.160	0.507	0.553	0.861	0.515	0.445	0.855	1.000	
Out-State Cites	-0.082	0.011	0.035	0.089	0.236	0.258	0.404	0.240	0.184	0.468	0.450	1.000

Notes. Observation is a judge-year, N=15,486. These statistics are constructed from each judge's yearly output of cases. Citation measures are per opinion. See variable definitions in the accompanying text.

TABLE 3.1
Effect of Reform: Establishment of Intermediate Appellate Court (26 States)

Outcome Variable	(1)	(2)	(3)
Log Number of Majority Opinions Written	-0.380**	-0.257**	-0.222**
	(0.1000)	(0.0834)	(0.0725)
Log Number of Dissents Written	-0.231+	-0.125	-0.107+
	(0.132)	(0.0828)	(0.0621)
Log Number of Concurrences Written	-0.128	-0.0464	-0.0345
	(0.111)	(0.0686)	(0.0627)
Log Total Words Written	-0.263**	-0.159*	-0.09
	(0.0844)	(0.0675)	(0.0557)
Log Average Length of Majority Opinion	0.116+	0.0959*	0.134**
	(0.0627)	(0.0416)	(0.0300)
Log Average Length of Table of Cases	0.084	0.0201	0.115**
	(0.0920)	(0.0579)	(0.0383)
Log Positive Cites Per Opinion	0.0239	0.0344	0.0858*
	(0.101)	(0.0481)	(0.0392)
Log Distinguishing Cites Per Opinion	0.169	0.110+	0.112*
	(0.124)	(0.0650)	(0.0536)
Log Negative Cites Per Opinion	0.129	0.119*	0.100*
	(0.0789)	(0.0480)	(0.0415)
Log Discuss Cites Per Opinion	-0.0173	0.0442+	0.0637**
	(0.0612)	(0.0257)	(0.0237)
Log Quoted Cites Per Opinion	0.0762	0.0317	0.0652*
	(0.0859)	(0.0347)	(0.0287)
Log Out-of-State Cites Per Opinion	-0.0783	0.0143	0.0420*
	(0.0661)	(0.0263)	(0.0194)
Log Cases Overruled	-0.0248	-0.0141	-0.0527
	(0.0513)	(0.0333)	(0.0449)
Log Cases Superseded by Statute	-0.0156	-0.0226	0.00691
	(0.0735)	(0.0473)	(0.0416)
Fixed Effects	None	State	Judge
Trends	None	State	State

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Treatment dummies include an interaction with whether the court has partial discretion or fully mandatory review.

TABLE 3.2
Effect of Establishing an Intermediate Appellate Court by Discretion Level

Outcome Variable	Baseline Effect (26 states)	Partial Discretion (11 states)	Mandatory Review (3 states)
Log Number of Majority Opinions Written	-0.222**	0.0954	0.164
	(0.0725)	(0.116)	(0.207)
Log Number of Dissents Written	-0.107+	0.0158	0.00994
	(0.0621)	(0.0719)	(0.144)
Log Number of Concurrences Written	-0.0345	-0.0171	-0.161*
	(0.0627)	(0.0705)	(0.0730)
Log Total Words Written	-0.09	-0.0121	0.152
	(0.0557)	(0.0807)	(0.104)
Log Average Length of Majority Opinion	0.134**	-0.109*	-0.00518
	(0.0300)	(0.0522)	(0.134)
Log Average Length of Table of Cases	0.115**	-0.0455	0.0696
	(0.0383)	(0.0534)	(0.127)
Log Positive Cites Per Opinion	0.0858*	-0.159*	0.0431
	(0.0392)	(0.0625)	(0.108)
Log Distinguishing Cites Per Opinion	0.112*	-0.0629	-0.102
	(0.0536)	(0.0653)	(0.120)
Log Negative Cites Per Opinion	0.100*	-0.078	-0.143+
	(0.0415)	(0.0475)	(0.0741)
Log Discuss Cites Per Opinion	0.0637**	-0.0970*	0.0432
	(0.0237)	(0.0400)	(0.0880)
Log Quoted Cites Per Opinion	0.0652*	-0.0806*	0.0463
	(0.0287)	(0.0389)	(0.105)
Log Out-of-State Cites Per Opinion	0.0420*	-0.115	0.0193
	(0.0194)	(0.0749)	(0.0687)
Log Cases Overruled	-0.0527	-0.0176	-0.0109
	(0.0449)	(0.0561)	(0.0906)
Log Cases Superseded by Statute	0.00691	-0.0509	0.0994
	(0.0416)	(0.0652)	(0.0618)

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each row includes coefficients from a regression that includes a year fixed effect, judge fixed effect, and state time trend. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Partial Discretion gives the estimate for the treatment dummy interacted with a dummy for states with partial discretion in case selection. Mandatory Review gives the estimate for the treatment dummy interacted with a dummy for states with fully mandatory review. Therefore the second and third column give effects that are relative to the states in the first column.

TABLE 3.3

Effect of Establishing an Intermediate Appellate Court on Bottom-Quality and Top-Quality Cases by Discretion Level

	Baseline Effect (26 states)		Partial Discretion (11 states)		Mandatory Review (3 states)	
	Bottom 5 Cases	Top 5 Cases	Bottom 5 Cases	Top 5 Cases	Bottom 5 Cases	Top 5 Cases
	(1)	(2)	(3)	(4)	(5)	(6)
Log Average Length of Majority Opinion	0.151**	0.0689**	-0.0661	-0.0904+	-0.0254	0.0704
	(0.0361)	(0.0228)	(0.0672)	(0.0454)	(0.162)	(0.0833)
Log Average Length of Table of Cases	0.158*	0.0616+	-0.00927	-0.0532	0.109	0.150*
	(0.0603)	(0.0355)	(0.0887)	(0.0321)	(0.173)	(0.0685)
Log Positive Cites Per Opinion	0.225**	-0.000212	-0.163*	-0.189+	0.0233	0.111+
	(0.0729)	(0.0304)	(0.0790)	(0.103)	(0.218)	(0.0583)
Log Distinguishing Cites Per Opinion	0.0455+	-0.00716	-0.0292	0.0246	-0.102*	-0.0212
	(0.0242)	(0.0494)	(0.0250)	(0.0703)	(0.0415)	(0.0913)
Log Negative Cites Per Opinion	0.0418*	0.0559	-0.0464*	-0.0696	-0.0802**	-0.129
	(0.0173)	(0.0401)	(0.0193)	(0.0528)	(0.0296)	(0.0930)
Log Discuss Cites Per Opinion	0.111**	-0.00428	-0.062	-0.132+	0.0534	0.130*
	(0.0375)	(0.0310)	(0.0478)	(0.0744)	(0.104)	(0.0578)
Log Quoted Cites Per Opinion	0.101*	-0.0193	-0.0695+	-0.124	0.056	0.128*
	(0.0409)	(0.0235)	(0.0414)	(0.0837)	(0.124)	(0.0558)
Log Out-of-State Cites Per Opinion	0.0496	-0.0202	-0.0306	-0.161	-0.022	0.155**
	(0.0346)	(0.0368)	(0.0414)	(0.134)	(0.0800)	(0.0533)
Log Cases Overruled	0.0359*	-0.0111	-0.0467*	0.00896	-0.0103	-0.0174
	(0.0139)	(0.0219)	(0.0188)	(0.0384)	(0.0213)	(0.0665)
Log Cases Superseded by Statute	0.00542	0.0056	-0.00978	-0.0617	0.0101	0.131**
	(0.00798)	(0.0231)	(0.00793)	(0.0401)	(0.0120)	(0.0430)

N= 13,438 judge-years. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Sample restricted to judge-years with at least 10 published cases. Sample of cases includes the bottom 5 or top 5 cases for each judge by year, as ranked by the number of positive citations. Each row includes coefficients from a regression that includes a year fixed effect, judge fixed effect, and state time trend. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline include the ten years before and after the policy change. Partial Discretion gives the estimate for the treatment dummy interacted with a dummy for states with partial discretion in case selection. Mandatory Review gives the estimate for the dummy interacted with a dummy for states with fully mandatory review. Therefore the second and third pairs of columns give effects that are relative to the states in the first pair of columns.

TABLE 4.1
Effect of Log Real Salary (All States)

Outcome Variable	(1)	(2)	(3)
Log Number of Majority Opinions Written	-0.588+	0.00518	-0.0513
	(0.310)	(0.285)	(0.286)
Log Number of Dissents Written	0.264	-0.585*	-0.195
	(0.331)	(0.220)	(0.190)
Log Number of Concurrences Written	0.236	-0.215	-0.0738
	(0.302)	(0.224)	(0.241)
Log Total Words Written	0.165	-0.0615	-0.0892
	(0.333)	(0.275)	(0.297)
Log Average Length of Majority Opinion	0.617*	0.0266	-0.0235
	(0.298)	(0.109)	(0.121)
Log Average Length of Table of Cases	0.428	0.154	0.0887
	(0.389)	(0.145)	(0.132)
Log Positive Cites Per Opinion	1.408**	0.336*	0.319+
	(0.477)	(0.153)	(0.182)
Log Distinguishing Cites Per Opinion	1.307*	0.198	0.219
	(0.514)	(0.155)	(0.175)
Log Negative Cites Per Opinion	0.960**	0.124	0.142
	(0.268)	(0.115)	(0.120)
Log Discuss Cites Per Opinion	1.030**	0.308*	0.286+
	(0.338)	(0.143)	(0.162)
Log Quoted Cites Per Opinion	1.255**	0.295+	0.292+
	(0.411)	(0.160)	(0.174)
Log Out-of-State Cites Per Opinion	0.319	0.107	0.0437
	(0.396)	(0.129)	(0.129)
Log Cases Overruled	-0.0639	-0.245	-0.407
	(0.271)	(0.324)	(0.353)
Log Cases Superseded by Statute	0.721*	0.0356	0.0472
	(0.353)	(0.223)	(0.248)
Fixed Effects	None	State	Judge
Trends	None	State	State

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a year fixed effect and indicators for all other treatment variables listed in Table 1.1. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Treatment dummies include an interaction with whether the court has partial discretion or fully mandatory review.

TABLE 4.2
Effect of Log Real Salary by Level of Review Discretion

Outcome Variable	Baseline Effect (50 states)	Partial Discretion (19 states)	Mandatory Review (9 states)
Log Number of Majority Opinions Written	-0.0513	0.0868	0.135
	(0.286)	(0.324)	(0.306)
Log Number of Dissents Written	-0.195	-0.273	0.161
	(0.190)	(0.254)	(0.228)
Log Number of Concurrences Written	-0.0738	0.0974	0.166
	(0.241)	(0.260)	(0.333)
Log Total Words Written	-0.0892	0.102	0.0889
	(0.297)	(0.279)	(0.446)
Log Average Length of Majority Opinion	-0.0235	0.0135	-0.0977
	(0.121)	(0.172)	(0.308)
Log Average Length of Table of Cases	0.0887	0.00739	-0.185
	(0.132)	(0.168)	(0.302)
Log Positive Cites Per Opinion	0.319+	-0.342	-0.963**
	(0.182)	(0.250)	(0.318)
Log Distinguishing Cites Per Opinion	0.219	-0.151	-0.264
	(0.175)	(0.208)	(0.222)
Log Negative Cites Per Opinion	0.142	-0.0948	-0.265*
	(0.120)	(0.118)	(0.115)
Log Discuss Cites Per Opinion	0.286+	-0.343+	-0.801*
	(0.162)	(0.199)	(0.301)
Log Quoted Cites Per Opinion	0.292+	-0.239	-0.747**
	(0.174)	(0.217)	(0.263)
Log Out-of-State Cites Per Opinion	0.0437	-0.152	-0.354
	(0.129)	(0.211)	(0.225)
Log Cases Overruled	-0.407	-0.0446	0.0106
	(0.353)	(0.349)	(0.387)
Log Cases Superseded by Statute	0.0472	0.0358	-0.113
	(0.248)	(0.290)	(0.296)

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each row includes coefficients from a regression that includes a year fixed effect, judge fixed effect, state time trend, and indicators for all other treatment variables listed in Table 1.1. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Partial Discretion gives the estimate for the treatment dummy interacted with a dummy for states with partial discretion in case selection. Mandatory Review gives the estimate for the treatment dummy interacted with a dummy for states with fully mandatory review. Therefore the second and third column give effects that are relative to the states in the first column.

TABLE 5.1
Effect of Reform: Term Length Increase (10 States)

Outcome Variable	(1)	(2)	(3)
Log Number of Majority Opinions Written	-0.119	-0.165	-0.132
	(0.305)	(0.219)	(0.227)
Log Number of Dissents Written	-0.56	-0.286	-0.163
	(0.401)	(0.237)	(0.211)
Log Number of Concurrences Written	-0.647+	-0.293+	-0.154
	(0.355)	(0.168)	(0.136)
Log Total Words Written	-0.181	-0.177	-0.106
	(0.250)	(0.165)	(0.197)
Log Average Length of Majority Opinion	-0.0227	-0.0267	-0.00917
	(0.0992)	(0.0916)	(0.0554)
Log Average Length of Table of Cases	-0.0122	0.0467	0.123**
	(0.153)	(0.0854)	(0.0349)
Log Positive Cites Per Opinion	-0.359	0.116*	0.138**
	(0.267)	(0.0496)	(0.0468)
Log Distinguishing Cites Per Opinion	0.219	0.212**	0.197**
	(0.233)	(0.0750)	(0.0620)
Log Negative Cites Per Opinion	0.0848	0.131**	0.124**
	(0.105)	(0.0460)	(0.0403)
Log Discuss Cites Per Opinion	-0.208	0.111**	0.105**
	(0.160)	(0.0351)	(0.0189)
Log Quoted Cites Per Opinion	-0.0701	0.114*	0.125**
	(0.202)	(0.0555)	(0.0368)
Log Out-of-State Cites Per Opinion	0.0545	0.0504	0.0453+
	(0.152)	(0.0465)	(0.0266)
Log Cases Overruled	-0.0758	-0.0427	-0.0126
	(0.0454)	(0.0370)	(0.0600)
Log Cases Superseded by Statute	-0.0219	0.0146	0.00506
	(0.0972)	(0.0750)	(0.0729)
Fixed Effects	None	State	Judge
Trends	None	State	State

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Treatment dummies include an interaction with whether the court has partial discretion or fully mandatory review.

TABLE 5.2
Effect of Term Length Increase by Level of Review Discretion

Outcome Variable	Baseline Effect (10 states)	Partial Discretion (2 states)	Mandatory Review (2 states)
Log Number of Majority Opinions Written	-0.132	-0.0927	-0.0346
	(0.227)	(0.260)	(0.195)
Log Number of Dissents Written	-0.163	-0.236	0.0124
	(0.211)	(0.205)	(0.223)
Log Number of Concurrences Written	-0.154	-0.141	0.0977
	(0.136)	(0.147)	(0.114)
Log Total Words Written	-0.106	-0.0358	-0.138
	(0.197)	(0.243)	(0.183)
Log Average Length of Majority Opinion	-0.00917	0.155+	-0.0562
	(0.0554)	(0.0811)	(0.0622)
Log Average Length of Table of Cases	0.123**	0.0963	-0.196**
	(0.0349)	(0.0822)	(0.0722)
Log Positive Cites Per Opinion	0.138**	0.168+	-0.210**
	(0.0468)	(0.0878)	(0.0672)
Log Distinguishing Cites Per Opinion	0.197**	0.0367	-0.105
	(0.0620)	(0.106)	(0.101)
Log Negative Cites Per Opinion	0.124**	0.0686	-0.0824+
	(0.0403)	(0.0638)	(0.0488)
Log Discuss Cites Per Opinion	0.105**	0.0616	-0.107+
	(0.0189)	(0.0579)	(0.0540)
Log Quoted Cites Per Opinion	0.125**	0.151*	-0.102*
	(0.0368)	(0.0657)	(0.0431)
Log Out-of-State Cites Per Opinion	0.0453+	0.00675	-0.0197
	(0.0266)	(0.101)	(0.0529)
Log Cases Overruled	-0.0126	0.283**	-0.0458
	(0.0600)	(0.0889)	(0.0597)
Log Cases Superseded by Statute	0.00506	0.0874	-0.127
	(0.0729)	(0.0954)	(0.0761)

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each row includes coefficients from a regression that includes a year fixed effect, judge fixed effect, and state time trend. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Partial Discretion gives the estimate for the treatment dummy interacted with a dummy for states with partial discretion in case selection. Mandatory Review gives the estimate for the treatment dummy interacted with a dummy for states with fully mandatory review. Therefore the second and third column give effects that are relative to the states in the first column.

TABLE 6.1
Effect of Reform: Non-Partisan Contested Elections to Uncontested Elections (6 States)

Outcome Variable	(1)	(2)	(3)
Log Number of Majority Opinions Written	0.148	0.0603	-0.0631
	(0.0906)	(0.0700)	(0.0582)
Log Number of Dissents Written	0.476*	0.148+	0.159*
	(0.179)	(0.0821)	(0.0659)
Log Number of Concurrences Written	0.331*	-0.00789	-0.00547
	(0.139)	(0.0570)	(0.0574)
Log Total Words Written	-0.171*	0.0898	-0.0345
	(0.0751)	(0.0657)	(0.0535)
Log Average Length of Majority Opinion	-0.410**	0.00299	-0.0168
	(0.0844)	(0.0490)	(0.0325)
Log Average Length of Table of Cases	-0.248*	0.204**	0.232**
	(0.100)	(0.0457)	(0.0338)
Log Positive Cites Per Opinion	0.129	0.174**	0.177**
	(0.0793)	(0.0434)	(0.0373)
Log Distinguishing Cites Per Opinion	0.142	0.195**	0.143**
	(0.102)	(0.0429)	(0.0398)
Log Negative Cites Per Opinion	0.0474	0.0998**	0.0629**
	(0.0588)	(0.0254)	(0.0213)
Log Discuss Cites Per Opinion	0.0478	0.0781*	0.0994**
	(0.0569)	(0.0317)	(0.0288)
Log Quoted Cites Per Opinion	0.00515	0.139**	0.115**
	(0.0805)	(0.0400)	(0.0328)
Log Out-of-State Cites Per Opinion	-0.0783	0.0357	0.0418
	(0.0568)	(0.0259)	(0.0330)
Log Cases Overruled	-0.166*	-0.0728+	-0.105*
	(0.0794)	(0.0408)	(0.0476)
Log Cases Superseded by Statute	0.125*	0.159**	0.112*
	(0.0481)	(0.0415)	(0.0429)
Fixed Effects	None	State	Judge
Trends	None	State	State

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Treatment dummies include an interaction with whether the court has partial discretion or fully mandatory review.

TABLE 6.2
Effect of Non-Partisan to Uncontested Election Reform by Level of Review Discretion

Outcome Variable	Baseline Effect (6 states)	Partial Discretion (1 state)	Mandatory Review (4 states)
Log Number of Majority Opinions Written	-0.0631	-0.391	0.0336
	(0.0582)	(0.275)	(0.129)
Log Number of Dissents Written	0.159*	-0.323	-0.157+
	(0.0659)	(0.256)	(0.0908)
Log Number of Concurrences Written	-0.00547	-0.23	0.153
	(0.0574)	(0.165)	(0.120)
Log Total Words Written	-0.0345	-0.209	0.066
	(0.0535)	(0.229)	(0.0787)
Log Average Length of Majority Opinion	-0.0168	0.141+	0.0614
	(0.0325)	(0.0729)	(0.0864)
Log Average Length of Table of Cases	0.232**	0.128**	-0.149*
	(0.0338)	(0.0424)	(0.0647)
Log Positive Cites Per Opinion	0.177**	-0.0437	-0.104+
	(0.0373)	(0.0532)	(0.0529)
Log Distinguishing Cites Per Opinion	0.143**	0.256**	0.127
	(0.0398)	(0.0761)	(0.0767)
Log Negative Cites Per Opinion	0.0629**	0.218**	0.0664
	(0.0213)	(0.0471)	(0.0436)
Log Discuss Cites Per Opinion	0.0994**	0.0345	-0.0218
	(0.0288)	(0.0376)	(0.0445)
Log Quoted Cites Per Opinion	0.115**	0.0917	-0.0465
	(0.0328)	(0.0587)	(0.0547)
Log Out-of-State Cites Per Opinion	0.0418	0.0841	-0.00112
	(0.0330)	(0.0525)	(0.0424)
Log Cases Overruled	-0.105*	-0.0216	-0.0987
	(0.0476)	(0.0837)	(0.0965)
Log Cases Superseded by Statute	0.112*	0.0481	0.170*
	(0.0429)	(0.0840)	(0.0754)

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each row includes coefficients from a regression that includes a year fixed effect, judge fixed effect, and state time trend. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Partial Discretion gives the estimate for the treatment dummy interacted with a dummy for states with partial discretion in case selection. Mandatory Review gives the estimate for the treatment dummy interacted with a dummy for states with fully mandatory review. Therefore the second and third column give effects that are relative to the states in the first column.

TABLE 7.1
Effect of Reform: Partisan Contested Elections to Uncontested Elections (9 States)

Outcome Variable	(1)	(2)	(3)
Log Number of Majority Opinions Written	0.164	0.0542	-0.0113
	(0.153)	(0.121)	(0.144)
Log Number of Dissents Written	0.224	0.00786	-0.0646
	(0.158)	(0.145)	(0.109)
Log Number of Concurrences Written	0.280*	0.11	0.0261
	(0.104)	(0.0742)	(0.0530)
Log Total Words Written	0.0679	-0.0115	-0.0963
	(0.0976)	(0.101)	(0.123)
Log Average Length of Majority Opinion	-0.0831	-0.0343	-0.0431
	(0.123)	(0.0605)	(0.0396)
Log Average Length of Table of Cases	-0.0487	-0.0277	-0.087
	(0.202)	(0.0923)	(0.0569)
Log Positive Cites Per Opinion	0.203	-0.0322	-0.117*
	(0.220)	(0.0737)	(0.0475)
Log Distinguishing Cites Per Opinion	-0.073	-0.0765	-0.112+
	(0.219)	(0.0945)	(0.0638)
Log Negative Cites Per Opinion	-0.0914	-0.0659	-0.0727
	(0.125)	(0.0596)	(0.0490)
Log Discuss Cites Per Opinion	0.131	-0.0303	-0.0611*
	(0.147)	(0.0368)	(0.0262)
Log Quoted Cites Per Opinion	0.00445	-0.0136	-0.0608+
	(0.191)	(0.0543)	(0.0344)
Log Out-of-State Cites Per Opinion	0.106	-0.0259	-0.0456*
	(0.114)	(0.0336)	(0.0178)
Log Cases Overruled	0.0555	0.0289	0.0182
	(0.0775)	(0.0305)	(0.0338)
Log Cases Superseded by Statute	-0.177+	-0.0574	-0.0812
	(0.0964)	(0.0532)	(0.0501)
Fixed Effects	None	State	Judge
Trends	None	State	State

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Treatment dummies include an interaction with whether the court has partial discretion or fully mandatory review.

TABLE 7.2
Effect of Partisan to Uncontested Election Reform by Level of Review Discretion

Outcome Variable	Baseline Effect (9 states)	Partial Discretion (4 states)	Mandatory Review (0 states)
Log Number of Majority Opinions Written	-0.0113	0.0866	
	(0.144)	(0.133)	
Log Number of Dissents Written	-0.0646	0.227+	
	(0.109)	(0.131)	
Log Number of Concurrences Written	0.0261	0.0468	
	(0.0530)	(0.116)	
Log Total Words Written	-0.0963	0.147	
	(0.123)	(0.112)	
Log Average Length of Majority Opinion	-0.0431	0.0162	
	(0.0396)	(0.0503)	
Log Average Length of Table of Cases	-0.087	0.143*	
	(0.0569)	(0.0654)	
Log Positive Cites Per Opinion	-0.117*	0.156	
	(0.0475)	(0.111)	
Log Distinguishing Cites Per Opinion	-0.112+	0.0872	
	(0.0638)	(0.0872)	
Log Negative Cites Per Opinion	-0.0727	0.0752	
	(0.0490)	(0.0546)	
Log Discuss Cites Per Opinion	-0.0611*	0.0593	
	(0.0262)	(0.0575)	
Log Quoted Cites Per Opinion	-0.0608+	0.0299	
	(0.0344)	(0.0487)	
Log Out-of-State Cites Per Opinion	-0.0456*	0.0584	
	(0.0178)	(0.0365)	
Log Cases Overruled	0.0182	0.0821	
	(0.0338)	(0.0687)	
Log Cases Superseded by Statute	-0.0812	0.122*	
	(0.0501)	(0.0578)	

N= 15,486 judge-years. Standard errors clustered by state in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each row includes coefficients from a regression that includes a year fixed effect, judge fixed effect, and state time trend. The baseline effect is a dummy for the ten years after a policy change, and the effect is relative to a baseline that includes the ten years before and after the policy change. Partial Discretion gives the estimate for the treatment dummy interacted with a dummy for states with partial discretion in case selection. Mandatory Review gives the estimate for the treatment dummy interacted with a dummy for states with fully mandatory review. Therefore the second and third column give effects that are relative to the states in the first column.

TABLE 8.1
Effect of Being up for Election By Election System

	State Fixed Effects and State Trends			Judge Fixed Effects and State Trends			State-Year Fixed Effects and Judge Fixed Effects		
	(NP1)	(P1)	(U1)	(NP2)	(P2)	(U2)	(NP3)	(P3)	(U3)
Log Number of Majority Opinions Written	-0.150*	-0.125**	0.198**	-0.227*	-0.153**	0.129*	-0.255+	-0.129**	0.091
	(0.0649)	(0.0242)	(0.0424)	(0.105)	(0.0255)	(0.0473)	(0.133)	(0.0162)	(0.0899)
Log Number of Dissents Written	-0.140+	-0.116**	0.0228	-0.151**	-0.0852	-0.0395	-0.148*	-0.081	-0.115
	(0.0693)	(0.0348)	(0.0212)	(0.0466)	(0.0581)	(0.0305)	(0.0597)	(0.0553)	(0.0836)
Log Number of Concurrences Written	-0.0669	-0.00757	0.111**	-0.0595*	0.0222	0.0349	-0.0477	0.0229	0.0547
	(0.120)	(0.0392)	(0.0392)	(0.0227)	(0.0280)	(0.0460)	(0.0431)	(0.0303)	(0.0601)
Log Total Words Written	-0.0784	-0.207**	0.138**	-0.191*	-0.217**	0.108**	-0.210*	-0.211**	0.0654
	(0.0495)	(0.0248)	(0.0342)	(0.0746)	(0.0407)	(0.0375)	(0.102)	(0.0259)	(0.0860)
Log Average Length of Majority Opinion	0.074	-0.03	-0.0379	0.0598	-0.0479*	-0.012	0.052	-0.0611**	0.00228
	(0.0613)	(0.0227)	(0.0494)	(0.0356)	(0.0205)	(0.0136)	(0.0358)	(0.0205)	(0.00350)
Log Average Length of Table of Cases	0.0341	-0.0425	0.0162	0.00985	-0.0556	0.0246+	0.0232	-0.0494	0.0231+
	(0.0679)	(0.0397)	(0.0333)	(0.0472)	(0.0405)	(0.0139)	(0.0471)	(0.0319)	(0.0138)
Log Positive Cites Per Opinion	0.0688	-0.00632	0.0379	0.0282	-0.0668**	0.0264	0.0386	-0.0617	-0.000557
	(0.0928)	(0.0298)	(0.0548)	(0.0428)	(0.0214)	(0.0323)	(0.0692)	(0.0428)	(0.0321)
Log Distinguishing Cites Per Opinion	0.0437	-0.0301	0.0307	0.0298	-0.0547+	0.03	0.000722	-0.0587**	0.00588
	(0.0693)	(0.0193)	(0.0856)	(0.0530)	(0.0273)	(0.0500)	(0.0439)	(0.0213)	(0.0464)
Log Negative Cites Per Opinion	0.0212**	0.000577	0.0113	0.0234*	-0.0116	0.0169	0.0330**	-0.0297	-0.00336
	(0.00678)	(0.0147)	(0.0349)	(0.0106)	(0.0184)	(0.0224)	(0.0125)	(0.0197)	(0.0262)

Notes. N= 9,502 judge-years. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. See notes in Table 8.1 (cont.).

TABLE 8.1 (cont.)
Effect of Being up for Election By Election System (cont.)

	State Fixed Effects and State Trends			Judge Fixed Effects and State Trends			State-Year Fixed Effects and Judge Fixed Effects		
	(NP1)	(P1)	(U1)	(NP2)	(P2)	(U2)	(NP3)	(P3)	(U3)
Log Discuss Cites Per Opinion	0.032	-0.0452**	0.0339	0.0172	-0.0716**	0.0252	0.0211	-0.0751*	0.0173
	(0.0612)	(0.0117)	(0.0461)	(0.0345)	(0.0197)	(0.0411)	(0.0502)	(0.0300)	(0.0433)
Log Quoted Cites Per Opinion	0.0461	-0.0269	0.0149	0.0304	-0.0646*	0.0155	0.0312	-0.0546**	-0.0155
	(0.0756)	(0.0183)	(0.0486)	(0.0383)	(0.0236)	(0.0244)	(0.0504)	(0.0210)	(0.0287)
Log Out-of-State Cites Per Opinion	0.0557	-0.0412*	0.0287	0.0585*	-0.0402*	0.0362	0.0224	-0.0361+	0.0534+
	(0.0410)	(0.0196)	(0.0354)	(0.0227)	(0.0163)	(0.0251)	(0.0489)	(0.0204)	(0.0278)
Log Cases Overruled	-0.08	-0.00105	0.207	-0.122	-0.0077	0.211	-0.0474	0.0101	0.207
	(0.0724)	(0.0282)	(0.156)	(0.0873)	(0.0305)	(0.167)	(0.0852)	(0.0209)	(0.260)
Log Cases Superseded by Statute	0.0151	-0.0364	0.126*	-0.0292	-0.0587	0.110**	-0.0538	-0.03	0.014
	(0.0218)	(0.0340)	(0.0469)	(0.0381)	(0.0426)	(0.0329)	(0.0797)	(0.0574)	(0.0329)

Notes. N= 9,502 judge-years. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a year fixed effect. The treatment variable is a dummy that equals one in years that a judge is facing re-election and zero otherwise. Regression contains interaction terms between treatment and level of review discretion. Sample is restricted to state-years with non-partisan (NP), partisan (P), and uncontested (U) elections, where the courts follow random or rotating assignment of cases. The sample includes 104 non-partisan elections for 66 judges in 9 states, 208 partisan elections for 114 judges in 10 states, and 77 uncontested elections for 51 judges in 7 states. State supreme court elections take place in the first week of November.

TABLE 8.2
Electoral Cycle Effect by Election System and Appellate Review Standard

	Non-Partisan Contested Elections			Partisan Contested Elections			Uncontested Elections		
Outcome Variable	Baseline Effect	Partial Discretion	Mandatory Review	Baseline Effect	Partial Discretion	Mandatory Review	Baseline Effect	Partial Discretion	Mandatory Review
Log Number of Majority Opinions Written	-0.255+	0.430**	0.225	-0.129**	0.131		0.091	0.0801	-0.398*
	(0.133)	(0.133)	(0.153)	(0.0162)	(0.137)		(0.0899)	(0.142)	(0.182)
Log Number of Dissents Written	-0.148*	0.0735	0.141	-0.081	0.132		-0.115	0.212+	-0.134
	(0.0597)	(0.223)	(0.227)	(0.0553)	(0.165)		(0.0836)	(0.117)	(0.373)
Log Number of Concurrences Written	-0.0477	-0.0813	0.115*	0.0229	-0.267**		0.0547	0.207	-0.367**
	(0.0431)	(0.114)	(0.0541)	(0.0303)	(0.0842)		(0.0601)	(0.153)	(0.0637)
Log Total Words Written	-0.210*	0.316**	0.191	-0.211**	0.174		0.0654	0.0771	-0.461
	(0.102)	(0.104)	(0.130)	(0.0259)	(0.169)		(0.0860)	(0.0876)	(0.365)
Log Average Length of Majority Opinion	0.052	-0.105**	-0.0308	-0.0611**	0.0213		0.00228	-0.0929	-0.0522
	(0.0358)	(0.0375)	(0.0456)	(0.0205)	(0.0305)		(0.00350)	(0.131)	(0.0881)
Log Average Length of Table of Cases	0.0232	-0.113*	0.0361	-0.0494	0.0157		0.0231+	-0.154	-0.042
	(0.0471)	(0.0485)	(0.0591)	(0.0319)	(0.0606)		(0.0138)	(0.0965)	(0.0997)
Log Positive Cites Per Opinion	0.0386	-0.200*	-0.0203	-0.0617	0.0275		-0.000557	-0.128	-0.024
	(0.0692)	(0.0947)	(0.0736)	(0.0428)	(0.0479)		(0.0321)	(0.231)	(0.0576)
Log Distinguishing Cites Per Opinion	0.000722	-0.0151	0.114+	-0.0587**	0.0352		0.00588	-0.303	-0.0913
	(0.0439)	(0.0804)	(0.0593)	(0.0213)	(0.0334)		(0.0464)	(0.314)	(0.155)
Log Negative Cites Per Opinion	0.0330**	-0.0841**	-0.0323*	-0.0297	0.0142		-0.00336	-0.141	-0.0513+
	(0.0125)	(0.0137)	(0.0139)	(0.0197)	(0.0210)		(0.0262)	(0.152)	(0.0262)

Notes. N= 9,502 judge-years. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. See notes in Table 8.2 (cont.).

TABLE 8.2 (cont.)
Electoral Cycle Effect by Election System and Appellate Review Standard (cont.)

	Non-Partisan Contested Elections			Partisan Contested Elections			Uncontested Elections		
Outcome Variable	Baseline Effect	Partial Discretion	Mandatory Review	Baseline Effect	Partial Discretion	Mandatory Review	Baseline Effect	Partial Discretion	Mandatory Review
Log Discuss Cites Per Opinion	0.0211	-0.0892	0.00984	-0.0751*	0.0714*		0.0173	-0.0402	-0.0606
	(0.0502)	(0.0601)	(0.0584)	(0.0300)	(0.0311)		(0.0433)	(0.0871)	(0.0435)
Log Quoted Cites Per Opinion	0.0312	-0.156*	0.00233	-0.0546**	0.0349		-0.0155	-0.0891	0.0995**
	(0.0504)	(0.0734)	(0.0580)	(0.0210)	(0.0257)		(0.0287)	(0.184)	(0.0287)
Log Out-of-State Cites Per Opinion	0.0224	-0.04	0.025	-0.0361+	0.0373		0.0534+	-0.148	-0.225
	(0.0489)	(0.0643)	(0.0507)	(0.0204)	(0.0438)		(0.0278)	(0.242)	(0.139)
Log Cases Overruled	-0.0474	-0.14	-0.0303	0.0101	0.0614+		0.207	-0.45	-0.112
	(0.0852)	(0.113)	(0.101)	(0.0209)	(0.0323)		(0.260)	(0.572)	(0.311)
Log Cases Superseded by Statute	-0.0538	0.143	0.106	-0.03	0.121		0.014	-0.231	-0.131
	(0.0797)	(0.168)	(0.107)	(0.0574)	(0.0849)		(0.0329)	(0.216)	(0.0949)

N= 12,405 judge-years. Standard errors in parentheses. + p < .1, * p < .05, ** p < .01, *** p < .001. Each coefficient is from a separate regression that includes a state-year fixed effect and judge fixed effect. The treatment variable is a dummy that equals one in years that a judge is facing re-election and zero otherwise. Regression contains interaction terms between treatment and level of review discretion. Sample is restricted to state-years with non-partisan (NP), partisan (P), and uncontested (U) elections, where the courts follow random or rotating assignment of cases. The sample includes 104 non-partisan elections for 66 judges in 9 states, 208 partisan elections for 114 judges in 10 states, and 77 uncontested elections for 51 judges in 7 states. State supreme court elections take place in the first week of November.